INTRODUCCIÓN A LA DOCUMÁTICA

1. Teoría



Jesús Tramullas

Página en blanco

Introducción a la Documática

1. Teoría

Jesús Tramullas Saz

Dep. Legal Z-3836-1996
ISBN 84-88502-42-7

La primera edición en soporte papel (1996) fue editada e impresa por Kronos S.L., Menéndez Pelayo, 4, 50009 Zaragoza

Página en blanco

Presentación (1997)

El impacto de las tecnologías de la información en las Ciencias de la Documentación (entendiendo este concepto como integrador de todas ellas) resulta ser de tal calibre en los últimos años, que lo que tradicionalmente se ha entendido como una técnica auxiliar está desplazando progresivamente a los métodos tradicionales, a los que sustituye con ventaja. Este desplazamiento es visto con preocupación en algunos sectores, que lo consideran un peligro, tendente a la desaparición de los especialistas en información y documentación como consecuencia de ésta, para ellos, excesiva "tecnologización". Sin embargo, nada más lejano de la realidad. Este panorama dominado por herramientas informáticas, si algo está poniendo de relieve, es que las tareas clave de tratamiento y recuperación de información necesitan de expertos y profesionales formados específicamente, tanto a nivel informático, como a nivel informativo y documental.

Formación que no puede circunscribirse al tratamiento de textos o a la creación de sencillas bases de datos: la utilización de interfaces gráficas de usuario, , la proliferación de sistemas de gestión documental administrativa, el auge de los hipermedia y de las bases de datos multimedia, de los sistemas expertos y de los sistemas basados en el conocimiento, y más recientemente el avance imparable de las llamadas autopistas de la información, obligan a mantener al día al especialista en información de manera casi frenética. Esfuerzo que debe desarrollarse tanto a nivel profesional como en los niveles educativos, responsabilidad de la Universidad.

Esta es la idea que inspira esta *Introducción a la Documática*, como ha sido bautizada en honor de P. Otlet. Ha sido cocnebida como una guía general de las principales informáticas necesarias para el experto en información y documentación, sin pretender ser exhaustiva, ya que en la actualidad puede encontrarse abundante bibliografía especializada en cada uno de los temas tratados, a la que el lector interesado puede recurrir para aumentar sus conocimientos. Su misión es servir como referencia, resumen, y, a la vez, punto de

acceso a las cuestiones de mayor importancia en el ámbito de la Documática.

Tras su redacción, si algo salta a la vista es la perentoria necesidad de replantear el concepto de documento. El documento actual puede encontrarse en muy diversos soportes, puede variar en tiempo y espacio, puede formarse utilizando partes componentes situadas en lugares distintos, y a veces sólo relacionadas por la voluntad de un autor. Puede ser dinámico, puede incluir no sólo información, sino también acciones a desarrollar automáticamente... Todo un universo de posibilidades que los fundadores de la Documentación sólo pudieron intuir.

La Documática se configura como una disciplina con una notoria vertiente práctica. Esta práctica es acompañada por una notable labor investigadora, apreciable con una simple revisión de las revistas de mayor raigambre en el ámbito de las Ciencias de la Información. Investigación que no sólo se centra en las herramientas: también se ocupa de la vertiente humana y social de las mismas, de los comportamientos y respuestas de los usuarios, y de sus mecanismos de comprensión y cognición. Una correcta comprensión, y por ende aprovechamiento, de la Documática sólo se consigue mediante la comprensión de los sistemas, de los procesos y de las aplicaciones en las que se ven inmersas las herramientas informática. Sirvan de paradigma estas palabras de Sun Tzu en *El Arte de la Guerra* (s. IV a.C.):

“Pues si el príncipe esclarecido y el general competente derrotan al enemigo cada vez que pasan a la acción, si sus hazañas se salen fuera de lo común, es gracias a la información previa. Lo que se ha llamado información previa no puede obtenerse de los espíritus, ni de las divinidades, ni de la analogía con acontecimientos pasados, ni de los cálculos. Es necesario obtenerlo de hombrrs que conozcan la situación del enemigo.”

Notas de lectura (2002)

1. Validez de los contenidos

El lector debe considerar que el texto corresponde con el original del mismo nombre escrito y publicado a comienzos de 1997. Consecuentemente, el acelerado desarrollo de las tecnologías de la información y de Internet hace que algunos capítulos y apartados no reflejen el estado actual del conocimiento en ese tema. Aunque el contenido siga siendo válido, es necesario complementarlo con bibliografía más reciente.

Esta versión digital llega al lector gratuitamente por cortesía de los poseedores del D.L. y del ISBN de la obra: el autor, Jesús Tramullas Saz, y la editorial, Kronos S.L.

2. Sistema de exploración y lectura

Este documento digital se estructura, en un primer nivel, en secciones, que a su vez se dividen, en un segundo nivel, en subsecciones o apartados. Todas ellas corresponden con la misma secuencia que la del el original impreso. En la parte superior de las páginas encontrará opciones de navegación que llevan a la sección anterior o a la siguiente, al índice general y a las pautas de lectura (este documento). La primera página de cada sección ofrece el índice parcial de la misma. Cuando se accede a los apartados individuales, se muestra un menú lateral que permite navegar por la estructura de la sección en la que se encuentra el lector.

3. Los dragones...

En cada inicio de sección, el lector encontrará un párrafo de "Expedición Tercera, o los dragones de la probabilidad", un relato escrito por Stanislaw Lem, y recogido en su obra *Ciberíada* (Barcelona: Bruguera, 1986). Cada cual puede sacar las conclusiones que quiera sobre la ciencia y la ciencia ficción...

4. Lecturas complementarias

Es difícil seleccionar y recomendar lecturas para actualizar el contenido. Debería empezar por revisar los números del *Journal of the American Society for Information Science* (JASIS), de 1997, 1998, 1999 y 2000. También encontrará trabajos interesantes en las

publicaciones españolas, especialmente en las *Actas de las Jornadas Españolas de Documentación* que organiza FESABID, y en los volúmenes que publica el capítulo español de ISKO (International Society for Knowledge Oragnization). Resulta interesante el apartado "Artículos" de la revista *El Profesional de la Información*. Como monografías, son especialmente recomendables KORFHAGE, R, R., *Information Storage and Retrieval.* John Wiley, 1997; BAEZA-YATES, R. y RIBEIRO-NETO, B., *Modern Information Retrieval.* ACM Press/Addison-Wesley, 1999; MEADOW, C.T, *Text Information Retrieval Systems*. Academic Press (Library and Information Science Series), 1999. Para la recuperación de información en Internet, TRAMULLAS, J., y OLVERA, Mª.L., *Recuperación de Información en Internet*. Madrid: Ra-Ma, 2001.

Recapitulación (2010)

Han transcurrido catorce años desde la redacción original de I*ntroducción a la Documática. 1. Teoría,* en 1996. Publicada en papel en enero de 1997, durante este tiempo la obra original ha sido ampliamente utilizada como material para estudiar oposiciones, preparar asignaturas, impartir cursos, y muchos otros usos (seguramente algunos de ellos inconfesables). Agotada rápidamente en su versión papel, se publicó en formato HTML en el web en 2002. Desde ese momento hasta 2010 ha estado disponible en el dominio http://docunautica.com, generando un tráfico que ha ido oscilando alrededor de 1000 visitantes y entre 3000 y 4000 páginas vistas cada mes. Estuvo a punto de ser retirada en 2006, pero numerosas peticiones de usuarios me persuadieron de que aún resultaba de utilidad, y que merecía la pena mantenerla accesible.

El devenir de mi trabajo en la Universidad me ha ido haciendo posponer la idea de redactar una actualización, que he ido dilatando en el tiempo. Finalmente, y tras un plazo de alejamiento y reflexión, no voy a llevar a cabo esa posible modernización. Esta reedición en formato digital de 2010 va a ser la última versión. No tengo la disponbilidad necesaria ni de tiempo ni de recursos que me exigía para redactarla con la calidad y profundidad adecuadas. Otros temas de investigación, más cercanos a los usuarios, acaparan mi interés y dedicación. A beneficio de la comunidad, he decidido liberar el contenido bajo una licencia *GNU Free Documentation License* (http://www.gnu.org/copyleft/fdl.html).

El lector encontrará partes que mantienen plena validez, y que sólo necesitan ser completados con los útlimos avances, y capítulos que suenan a históricos, como el dedicado a la teledocumentación, o algunas partes sobre internet, junto a otros plenamente superados, como la parte dedicada al formato MARC. La acelerada evolución de la disciplina hace difícil tener una perspectiva histórica del tremendo cambio que ha supuesto la última década. En todo caso, y lberado el contenido, puede ser el momento de que otros autores colaboren en el desarrollo y evolución de la obra original. Quizá un wiki colaborativo, en plena efervescencia de la gestión de información 2.0, puede ser una buena idea...

La relectura de la *Presentación* escrita para la primera edición me ha mostrado que muchos de los conceptos e ideas que sirvieron para delinear el contenido y orientación de este libro siguen vigentes, o se han cumplido. El cambio en el concepto de documento, la interiorización de la tecnología en el trabajo de gestión de información, o la amenaza latente para el futuro del trabajo informativo-documental, sólo evitable a través de la formación continua y del cambio de actitud en la actividad profesional, han sido cuestiones que se han relevado como claves. Termino esta breve recapitulación igual que terminé la *Presentación* de la primera edición de 1997, citando como paradigma un fragmento de un libro, que en esta ocasión es un clásico del cyberpunk, *Neuromante*, escrito por W, Gibson:

"Operaba en un estado adrenalítico alto y casi permanente, conectado a una consola de ciberespacio hecha por encargo que proyectaba su incorpórea conciencia en la alucinación consensual que era la matriz."

1. HACIA UN CONCEPTO BÁSICO DE DOCUMENTACIÓN AUTOMÁTICA/DOCUMÁTICA.

Lecturas recomendadas: LÓPEZ YEPES, 1995; ESPINOSA *et al*, 1994; ELLIS, 1990; INGWERSEN, 1992; BLAIR, 1990; DEVLIN, 1991; DE MEY, 1980; BUCKLAND, 1991; CRONIN y VAKKARI, 1992; VICKERY y VICKERY, 1992.

1.1. P. Otlet: Tecnología y Documentación.

Las Ciencias de la Información y la Documentación, cuya formulación moderna es necesario buscar tomando como punto de partida las obras de P. Otlet, particularmente su monumental *Traité de Documentation*, publicado en 1934, han estado íntimamente relacionadas con el desarrollo de los medios técnicos y tecnológicos. Otlet, como han señalado otros trabajos (ESPINOSA *et al*., 1994) ya señaló en su momento a la tecnología como una disciplina general, amplia, que estudia los medios mediante los cuales el hombre intervenía en su medio ambiente. La Documentación es una disciplina que, correspondientemente, se encarga de intervenir en los procesos informativos desarrollados por el ser humano, y en la que interviene, con un papel fundamental, la tecnología, o aquella parte del conjunto total de la misma que es aplicable de forma útil al trabajo informativo.

De esta forma, Otlet expone los principios y relaciones de la Tecnología documental, insistiendo en un punto clave: que la tecnología documental debe elaborarse por especificación de la tecnología general. Es decir, que el desarrollo, métodos y técnicas de la Tecnología Documental dependerá de los desarrollos de la tecnología general. La rama de la Documentación encargada de la teorización, metodología y aplicación de los medios tecnológicas es identificada bajo los nombres de Bibliología Tecnológica o Documentación Técnica, mostrando una partición cruzada, dentro de la cual es posible identificar conceptos de plena actualidad.

En primer lugar, Otlet identifica los factores que caracterizan la nueva subdisciplina: existe una parte encargada de estudiar los soportes, y otra que se centra en las máquinas que hacen posible la misma: esta vertiente recibe el clarificador nombre de Documentación Automática. Además, estos medios se emplean en un proceso documental completo, desde la producción del documento hasta su difusión, con unos objetivos claros: ofrecer información actualizada, universal, rápida y verdadera, lo que en la actualidad puede englobarse bajo el término "optimizada". En segundo lugar, coloca en un lugar preponderante la importancia de la organización, como criterio indispensable en la Documentación, y por extensión a cualquier parte componente de la misma. Por último, aunque Otlet lo sitúa en su trabajo previamente, también presupone la existencia de una parte de la Documentación Automática encargada de estudiar el aspecto físico del documento, y de parte otra centrada en el estudio de los procesos documentales.

El breve resumen realizado permite identificar en la obra de Otlet los componentes fundamentales del moderno concepto de Documentación Automática (o Automatizada): una teoría sobre la organización, las herramientas y los soportes; una aplicación práctica al proceso documental, y, por último, un objetivo: satisfacer las necesidades informativas del usuario.

1.2. De Otlet al concepto de Information Retrieval.

El desarrollo posterior de la Documentación, así como el nacimiento de varias escuelas teóricas dentro del mismo ámbito, y el desarrollo de diferentes enfoques y perspectivas sobre la misma ha sido objeto de concienzudos estudios (LÓPEZ YEPES, 1995). Sin embargo, en todas ellas el aspecto tecnológico se ha considerado como un medio para desarrollar tareas, profundizando sólo en uno de los ámbitos propuestos por Otlet para la Documentación Automática.

Hay que esperar al desarrollo del concepto de *Information Retrieval*, especialmente en el ámbito anglosajón, en la década de 1950, para la aparición de un corpus teórico y experimental en el que las tecnologías de la información ocupan un destacado papel. Independientemente de la acuñación del término, es este entorno científico y social el que marca la aparición del concepto y el recurso continuo a las tecnologías de la información, esencialmente ordenadores y redes de comunicaciones. Las décadas de 1950 y 1960 ven el auge que alcanzan los estudios sobre comunicación, psicología, lógica, cibernética, telecomunicaciones, teoría de sistemas... al tiempo que una imparable multiplicación de los recursos informativos en todos los campos de la actividad humana. Los especialistas se centran en el problema de la búsqueda y recuperación de información, documental o no, lo que les lleva al comienzo del ciclo, a los principios de representación y clasificación de los documentos según su contenido semántico. *Information Retrieval* era un síntoma de un cambio más profundo, que llegó en 1966-68, con la introducción de la *Information Science* (Ciencia de la Información), como ciencia integradora de la teoría, proceso y práctica documental, con otras ciencias complementarias, como la cibernética, la informática, la teoría de la información y la comunicación, etc. Biblioteconomía y Documentación serían, desde esta perspectiva, aplicaciones particulares de la *Information Science*.

Del impacto de la informática (no debe olvidarse que en Norteamérica no se utiliza un término general como informática, sino que se utilizan sus partes como *computer science, computer engineering, data processing*...) se deduce la presencia de términos que definen a profesionales relacionados con la tecnología, como *information specialist* o *information engineer*. El éxito de la *Information Science*, independientemente de los problemas teóricos y conceptuales pertinentes a toda disciplina científica, facilitaron introducción cada vez más profunda de las tecnologías de la información, en especial la informática, en numerosos aspectos del proceso informativo documental. El impacto favorece que, en ámbito anglosajón se acuñe el término *Informatics* como la unión entre *Computer Sciences* e *Information Science*.

El desarrollo de nuevas teorías ha traído, de la mano de la Ciencia de la Información, la aparición de la disciplina *Information Management* (Gestión de Información y la Documentación en las Organizaciones), en la que desempeñan un papel fundamental las telecomunicaciones y la informática, íntimamente relacionadas con los sistemas de información, en el marco de redes complejas de información.

La escuela soviética que se acoge al nombre de *Informatika* también concede un papel importante a las tecnologías de la información. El problema derivado de la utilización del término *Informatika*, y la dificultad para su aceptación en el ámbito europeo, a causa de la homofonía con el término que engloba las ciencias de los ordenadores y sus aplicaciones (Informática), ha sido la causa de proponer un nuevo término, Informología, con un contenido similar, aunque desde un posicionamiento teórico original, a la Ciencia de la Información, incluyendo sus vertientes tecnológicas.

1.3. El enfoque cognitivo.

El desarrollo e implantación de la *Information Science* han sido claves en la madurez que ha alcanzado la aplicación de tecnologías de la información en el trabajo informativo. La utilización de medios informáticos ha corrido pareja con el desarrollo de nuevas máquinas y herramientas desde la década de 1960. Los diferentes estadios tecnológicos de los medios informáticos han tenido su reflejo en el tratamiento de la información, y, por ende, en los enfoques teóricos que han guiado la misma. De Mey ha trazado una útil evolución del desarrollo del tratamiento de información (DE MEY, 1980):

1. Una etapa monádica en la cual las unidades de información se tratan independientemente unas de otras, como diferentes entidades existentes por sí mismas (por ejemplo, un libro con un término que lo define).
2. Una etapa estructural, en la que la información se ve como una entidad compleja, compuesta de unidades simples de información, relacionadas de algún modo, cuya estructura es objeto de estudio (por ejemplo, las frases o párrafos en un libro, con palabras clave y análisis sintáctico)).
3. Una etapa contextual, en la que al enfoque de la etapa anterior se une un análisis del contexto en el que se sitúa la información, para eliminar o

reducir la ambigüedad que puede producirse en el significado del mensaje (por ejemplo, utilizar los valores semánticos)
4. Una etapa cognitiva, en la que la información se ve como un complemento a un sistema conceptual que representa el conocimiento del sistema de procesado de información del mundo (por ejemplo, sistemas basados en el conocimiento y sistemas adaptativos).

Para este autor, los actuales sistemas informáticos habrían cubierto completamente las fases 1 y 2, actuarían con apoyo humano en la fase 3, y se encontrarían en un momento experimental en la fase 4.

Esta fase 4 es el objeto principal de investigación de Ingwersen (INGWERSEN , 1992), y ha sido tratado por otros autores, entre los que debe destacarse Ellis (ELLIS, 1990). Para Ingwersen pueden diferenciarse varios enfoques en el tratamiento y recuperación de la información, coincidentes con el desarrollo histórico de la informática y de las fases propuestas por De Mey, que culminan con la utilización de un enfoque cognitivo. De esta forma diferencia un primer enfoque tradicional, centrado en los problemas de representación del texto y recuperación de información mediante técnicas mecanicistas. La superación del anterior se obtiene mediante un enfoque orientado al usuario, potenciando la figura del intermediario humano, y considerando todos los procesos de comunicación y representación de la información. El enfoque cognitivo, por último, pretende superar los problemas de representación y recuperación de la información mediante la creación de mecanismos que reflejan las estructuras cognitivas del usuario, tanto en lo que respecta a modelos como a estados. Puede deducirse, por lo tanto, que en este enfoque cobran especial importancia tanto la semántica del texto y el estudio del lenguaje natural, en lo explícito y en lo implícito, como en la contextualización, en su más amplia acepción, de las necesidades de información del usuario en un momento y situación dados. Para Ingwersen, el desarrollo de nuevos y mejores mecanismos de tratamiento y recuperación de la información debe basarse en la combinación de los mecanismos clásicos de recuperación de información, con métodos y elementos de polirepresentación, contextualización e interacción que son relevantes a manifestaciones de necesidades de información de los usuarios.

1.4. Hacia un concepto de Documentación Automática.

La construcción de un concepto en el ámbito científico tiene dos vertientes: en primer lugar, la elección del término que lo englobará, y en segundo lugar, la redacción de la definición del término. Problema éste presente en todas las ciencias, por el propio carácter dinámico de las mismas, siempre sometidas a continua revisión, al que no son ajenas las Ciencias de la Información y la Documentación, por su propia riqueza epistemológica. En el ámbito que nos ocupa la numerosa terminología, en ocasiones contradictoria, ha favorecido la aparición de términos que pretenden englobar al objeto de estudio: bibliología tecnológica, *informatics,* informática documental, *documentique, documatique, informatika,* documática, tecnologías de la información, *information retrieval*... cada una de las cuales con su definición correspondiente, en las que se pueden

identificar matices y particularidades de diversos tipos y niveles, aunque todos ellos ofrecen un sustrato común.

Ante esta situación, es necesario comenzar por el término. En esta vertiente resulta adecuado inclinarse, al menos desde nuestra perspectiva, por la vuelta a la fuente clásica, el trabajo de Otlet, citado con anterioridad. Sin entrar en disquisiciones sobre el término "automática", resulta evidente que Otlet ofreció con claridad meridiana un término válido en la actualidad, como es Documentación Automática, tomando como punto de partida la situación tecnológica del tiempo que le tocó vivir. La utilización del calificativo "Automática" resulta de suma importancia, ya que está haciendo referencia directa a la aplicación de determinadas tecnologías. Otros términos posteriores pueden dar lugar, por el problema terminológico citado, a cuestiones sobre la automatización o informatización de las tareas que definen (en puridad, puede hablarse de *information retrieval* sin tratar cuestiones informáticas). Desde otro punto de vista, distinguir entre Biblioteconomía Automática, Archivística Automática, Gestión de Información Automática, etc, etc, parece un ejercicio de división innecesario a un primer nivel, ya que comparten teoría, métodos y herramientas con la Documentación Automática, distinguiéndose sólo por los componentes contextual y organizativo. Por lo tanto, se prefiere el término Documentación Automática, independientemente de que los términos sean unidos para formular el nuevo término Documática.

Una vez adoptado el término de Documentación Automática/Documática, debe resolverse el problema de la definición, sus componentes y organización. Para establecer un *continuum* con el criterio de unidad adoptado en lo referente al término, la redacción de la definición debe ofrecer una integración de elementos restrictivos que permitan dotar al concepto de unos límites adecuados, con elementos sincréticos que doten a la definición de validez frente a nuevos desarrollos tecnológicos. Para ello se toman como punto de partida dos definiciones de *Information Retrieval* que son pertinentes:

> "Information retrieval is concerned with the processes involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desired by a human user... The objetive is to study and understand IR processes in order to design, build and test retrieval systems that may facilitate the effective communication of desired information between human generator and human user." (INGWERSEN, 1992: 49)

> "The central problem of Information Retrieval is how to represent documents for retrieval... The central task of Information Retrieval research is to understand how documents should be represented for effective retrieval." (BLAIR, 1990: vii).

En ambas definiciones se insiste en la importancia de la representación de documentos, con la finalidad de poder ofrecer respuestas adecuadas a una necesidad de información planteada. La definición de Ingwersen añade dos elementos más de sumo interés. En primer lugar, el diseño de sistemas para la

recuperación de información; en segundo lugar, el proceso de comunicación inherente a todo proceso informativo y documental. Sin embargo, ninguna de las dos definiciones hace referencia a los medios tecnológicos que pueden ser de aplicación, ni a criterios de organización relacionados con el contexto en el que se sitúe el sistema. Debe señalarse, además que ambas hacen referencia no a datos, sino a información (*information*) y a documentos (*documents*).

Esta última cuestión resulta de suma importancia en nuestro planteamiento, ya que permite diferenciar la disciplina en discusión de otras como el procesado de datos, al que, por norma general se adscribe a la disciplina informática. Esta sutil diferencia, ya señalada por Blair (BLAIR, 1990) va a servir como punto de partida para un análisis del material de trabajo de Biblioteconomía, Archivística y Documentación. A la pregunta de cuál es el material de trabajo de un profesional de estos ámbitos, la respuesta más rápida es "libros, seriadas, protocolos, documentación administrativa..." Es necesario realizar un análisis previo: cuando un profesional se enfrenta a un libro o a un protocolo, lo analiza, establece sus características físicas, de forma acorde a las normas existentes, para posteriormente realizar una representación de su contenido informativo (mediante la aplicación de lenguajes documentales u otro tipo de métodos). Ahí termina el trabajo con el documento primario. Además, ese trabajo ha consistido en extraer información, que no datos, para generar una representación plenamente informativa del documento primario. A partir de ese momento, todas las acciones del profesional, independientemente del tipo de unidad de información en el que se encuentre, se centran en la información que generó, que utiliza para satisfacer necesidades de usuarios y para los mecanismos de gestión y planificación de la unidad. Sólo se recurre al documento original para su préstamo o consulta temporal por parte del usuario. Esta diferencia es clave, a nuestro juicio, para diferenciar la Documentación Automática dentro del ámbito general de la disciplina Informática.

Desde estas consideraciones, la Documentación Automática /Documática es la disciplina que se ocupa de la investigación y aplicación de las Tecnologías de la Información en todos los ámbitos de las Ciencias de la Documentación, teniendo como material de trabajo la información generada y requerida durante el proceso documental que se trate, en el marco de los sistemas de información, en un contexto que integra los medios automáticos, el ser humano y las interacciones entre ambos.

La Documentación Automática/Documática ofrece dos vertientes; correspondientes a la investigación y desarrollo de nuevas aplicaciones, y a la utilización práctica de las mismas. En la primera participa de la interdisciplinariedad con otras ciencias, como la Informática, las Ciencias Cognitivas o las Ciencias de la Organización y la Gestión. El objetivo es obtener nuevas aplicaciones o herramientas informáticas para el tratamiento y recuperación de la información, así como métodos para la implantación, gestión y control de estas aplicaciones o herramientas dentro de diferentes tipos de organizaciones. La segunda es la encargada de aplicar las herramientas al entorno o contexto que se precise, de forma que se optimicen los procesos documentales e informativos que se desarrollen en el mismo. Ambas se encuentran relacionadas, ya que la investigación se beneficia y alimenta de las

experiencias obtenidas en la aplicación de las herramientas que desarrolla, y a su vez las nuevas herramientas permiten mejorar y adecuar cada vez más los procesos documentales que se desarrollan en las unidades de información, al mismo tiempo que intervienen, en numerosas ocasiones, en los procesos de organización y gestión de las mismas. Junto a esta interacción, no debe olvidarse la integración, tanto vertical como horizontal, de la Documentación Automática/Documática, tanto con la Ciencia de la Documentación, como con la Informática y la Ciencia de la Información, ésta última en su más amplia formulación.

Página en blanco

2. LOS SISTEMAS DE BASES DE DATOS Y LOS SGBD.

Lecturas recomendadas: ELMASRI y NAVATHE, 1989; MIGUEL y PIATTINI, 1993; FROST, 1989; KORTH y SILBERSCHATZ, 1993; MOTA, CELMA y CASAMAYOR, 1994; BATINI, CERI y NAVATHE, 1994; JACKSON, 1990; REINGRUBER y GREGORY, 1994, BERTINO y MARTINO, 1995.

Una de las conclusiones obtenidas en el capítulo anterior era que los recursos de información, y los mecanismos necesarios para su interrogación, resultaban ser uno de los objetivos fundamentales en los sistemas de información que debían hacer frente a grandes cantidades de documentos e información en diferentes formatos y soportes. Y uno de los componentes principales de estos sistemas son las bases de datos, o, más concretamente, los sistemas de bases de datos. Resultará evidente, por otra parte, que la gestión del gran volumen de datos demandará una teoría sobre la organización de esos datos para alcanzar la máxima efectividad posible. En este capítulo se va a efectuar una revisión de los principios que inspiran la arquitectura y el diseño de sistemas de bases de datos, poniendo especial interés en el modelo entidad-relación, por cuanto será el utilizado más adelante para el diseño de las bases de datos con sistemas de gestión de bases de datos relacionales, ya que gran parte de los métodos de modelado conceptual pueden utilizarse igualmente en la construcción de bases de datos documentales.

2.1. Las bases de datos.

En el entorno informático, la gestión de bases de datos ha evolucionado desde ser una aplicación más disponible para los computadores, a ocupar un lugar fundamental en los sistemas de información. En la actualidad, un sistema de información será más valioso cuanto de mayor calidad sea la base de datos que lo soporta, la cual resulta a su vez un componente fundamental del mismo, de tal

forma que puede llegarse a afirmar que es imposible la existencia de un sistema de información sin una base de datos, que cumple la función de "memoria", en todas sus acepciones posibles, del sistema.

Las bases de datos almacenan, como su nombre dice, datos. Estos datos son representaciones de sucesos y objetos, a diferente nivel, existentes en el mundo real: en su conjunto, representan algún tipo de entidad existente. En el mundo real se tiene percepción sobre las entidades u objetos y sobre los atributos de esos objetos; en el mundo de los datos, hay registros de eventos y datos de eventos. Además, en ambos escenarios se puede incluso distinguir una tercera faceta: aquella que comprende las definiciones de las entidades externas, o bien las definiciones de los registros y de los datos.

La transferencia entre las entidades del mundo real, y sus características, y los registros contenidos en una base de datos, correspondientes a esas entidades, se alcanza tras un proceso lógico de abstracción, conjunto de tareas que suelen englobarse bajo el título de diseño de bases de datos.

Sin embargo, es necesario definir, en primer lugar, qué es una base de datos, independientemente de su diseño y/o su orientación. Entre las numerosas definiciones que pueden encontrarse en la bibliografía, pueden escogerse, por su exhaustividad, las siguientes:

> "Colección de datos correspondientes a las diferentes perspectivas de un sistema de información (de una empresa o institución), existentes en algún soporte de tipo físico (normalmente de acceso directo), agrupados en una organización integrada y centralizada en la que figuran no sólo los datos en sí, sino también las relaciones existentes entre ellos, y de forma que se minimiza la redundancia y se maximiza la independencia de los datos de las aplicaciones que los requieren." (GUILERA, 1993: 377)

> "Una base de datos es una colección de datos estructurados según un modelo que refleje las relaciones y restricciones existentes en el mundo real. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción han de ser únicas estando almacenadas junto a los mismos. Por último, los tratamientos que sufran estos datos tendrán que conservar la integridad y seguridad de éstos." (MOTA, CELMA y CASAMAYOR, 1994: 9)

La segunda definición añade los objetivos que debe cumplir un sistema de gestión de bases de datos, sobre los cuales se tratará más adelante. Por ahora, basta considerar que deben cumplir los objetivos de independencia de los datos (las aplicaciones no deben verse afectadas por cambios en la estructura de los datos), integridad de los datos (los datos deben cumplir ciertas restricciones que aseguren la correcta introducción, modificación y borrado de los mismos) y seguridad (establecer diferentes niveles de acceso a los datos a diferentes tipos de usuarios).

La entidad existente en el mundo real es objeto de un doble tratamiento, desde el momento en que convierte en objeto de la base de datos. El tratamiento de sus datos se va a realizar en un nivel lógico, por una parte, y en un nivel físico, por otra. En el primero de ellos, el lógico, se va a trabajar en los aspectos referidos a la identificación de las características de la entidad, su descripción y organización, mientras que en el segundo todo lo anterior se va a plasmar en la organización, acceso y almacenamiento de los datos en un soporte físico. Esta división entre un nivel lógico y otro físico se va reflejar en todos los métodos y conceptos subsiguientes.

2.2. El modelo de arquitectura de bases de datos.

Hasta fecha relativamente cercana, las bases de datos eran el resultado de una compleja programación y de complicados mecanismos de almacenamiento. Con la popularización de la microinformática, la aparición de aplicaciones específicas también trajo con ella la disponibilidad de herramientas de gestión de datos, que acabaron desembocando en los denominados sistemas de gestión de bases de datos, identificados por sus siglas SGBD (DBMS en inglés, siglas de DataBase Management Systems). De esta manera, la gestión de base de datos pudo liberarse de los grandes ordenadores centrales, pudiendo distribuirse según los intereses de los usuarios, y dotando de autonomía en la gestión de información a muchas entidades. Los SGBD permitieron a todo tipo de usuarios crear y mantener sus bases de datos, dotándolos de una herramienta que era capaz de transformar el nivel lógico que éstos diseñaban en un conjunto de datos, representaciones y relaciones, traduciéndolo al nivel físico correspondiente. Para que fuese posible, y para asegurar a los usuarios cierta seguridad en el intercambio de datos entre diferentes sistemas, y en el diseño de ficheros y bases de datos, fue necesario normalizar los esquemas que guiaban la creación de las bases de datos.

Las bases de datos respetan la arquitectura de tres niveles definida, para cualquier tipo de base de datos, por el grupo ANSI/SPARC. En esta arquitectura la base de datos se divide en los niveles externo, conceptual e interno (KORTH y SILBERSCHATZ, 1994:5; MIGUEL y PIATTINI, 1993: 83-107; MOTA, CELMA y CASAMAYOR, 1994: 11-12):

1. Nivel interno: es el nivel más bajo de abstracción, y define cómo se almacenan los datos en el soporte físico, así como los métodos de acceso.
2. Nivel conceptual: es el nivel medio de abstracción. Se trata de la representación de los datos realizada por la organización, que recoge las vistas parciales de los requerimientos de los diferentes usuarios y las aplicaciones posibles. Se configura como visión organizativa total, e incluye la definición de datos y las relaciones entre ellos
3. Nivel externo: es el nivel de mayor abstracción. A este nivel corresponden las diferentes vistas parciales que tienen de la base de datos los diferentes usuarios. En cierto modo, es la parte del modelo conceptual a la que tienen acceso.

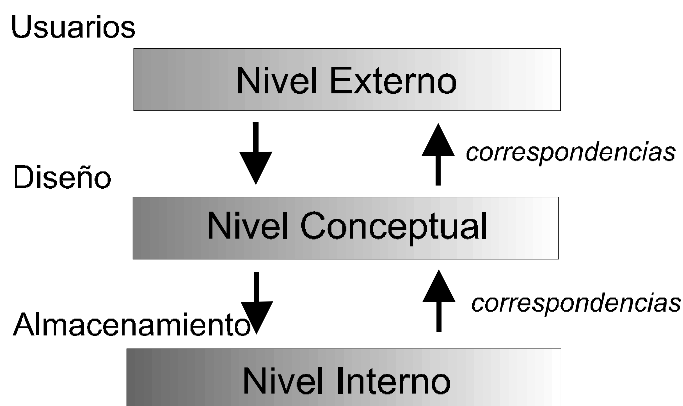


Fig.2.1. Niveles de la arquitectura de bases de datos.

En ocasiones puede encontrarse el nivel conceptual divido en dos niveles, conceptual y lógico. El primero de ellos corresponde a la visión del sistema global desde un punto de vista organizativo independiente, no informático. El segundo correspondería a la visión de la base de datos expresada en términos del sistema que se va a implantar con medios informáticos.

El modelo de arquitectura propuesto permite establecer el principio de independencia de los datos. Esta independencia puede ser lógica y física. Por independencia lógica se entiende que los cambios en el esquema lógico no deben afectar a los esquemas externos que no utilicen los datos modificados. Por independencia física se entiende que el esquema lógico no se vea afectado por cambios realizados en el esquema interno, correspondientes a modos de acceso, etc.

2.3. Los modelos de datos.

En el proceso de abstracción que conduce a la creación de una base de datos desempeña una función prioritaria el modelo de datos. El modelo de datos, como abstracción del universo de discurso, es el enfoque utilizado para la representación de las entidades y sus características dentro de la base de datos, y puede ser dividido en tres grandes tipos (KORTH y SILBERSCHATZ, 1993: 6-11):

1. Modelos lógicos basados en objetos: los dos más extendidos son el modelo entidad-relación y el orientado a objetos. El modelo entidad-relación (E-R) se basa en una percepción del mundo compuesta por objetos, llamados entidades, y relaciones entre ellos. Las entidades se diferencian unas de otras a través de atributos. El orientado a objetos también se basa en objetos, los cuales contienen valores y métodos,

entendidos como órdenes que actúan sobre los valores, en niveles de anidamiento. Los objetos se agrupan en clases, relacionándose mediante el envío de mensajes. Algunos autores definen estos modelos como "modelos semánticos".
2. Modelos lógicos basados en registros: el más extendido es el relacional, mientras que los otros dos existentes, jerárquico y de red, se encuentran en retroceso. Estos modelos se usan para especificar la estructura lógica global de la base de datos, estructurada en registros de formato fijo de varios tipos. El modelo relacional representa los datos y sus relaciones mediante tablas bidimensionales, que contienen datos tomados de los dominios correspondientes. El modelo de red está formado por colecciones de registros, relacionados mediante punteros o ligas en grafos arbitrarios. el modelo jerárquico es similar al de red, pero los registros se organizan como colecciones de árboles. Algunos autores definen estos modelos como "modelos de datos clásicos".
3. Modelos físicos de datos: muy poco usados, son el modelo unificador y el de memoria de elementos. Algunos autores definen estos modelos como "modelos de datos primitivos".

De lo anterior se deduce que el punto clave en la construcción de la base de datos será el modelo de datos. Se denomina modelo:

> "...al instrumento que se aplica a una parcela del mundo real (universo del discurso) para obtener una estructura de datos a la que denominamos esquema. Esta distinción entre el modelo (instrumento) y el esquema (resultado de aplicar el instrumento) es importante... Es importante también distinguir entre mundo real y universo del discurso, ya que este último es la visión que del mundo real tiene el diseñador... podemos definir un modelo de datos como un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos del universo del discurso." (MIGUEL y PIATTINI, 1993: 162)

Los objetivos del modelo de datos son dos:

1. Formalización: definir formalmente las estructuras permitidas y las restricciones a fin de representar los datos de un SI.
2. Diseño: el modelo resultante es un elemento básico para el desarrollo de la metodología de diseño de la base de datos.

Los diferentes modelos de datos comparten, aunque con diferentes nombres y notaciones, unos elementos comunes, componentes básicos de la representación de la realidad que realizan. Estos componentes se identifican gracias a la clasificación, y pueden identificarse conceptos estáticos y conceptos dinámicos. Los conceptos estáticos corresponden a:

1. Objeto: cualquier entidad con existencia independiente sobre el que almacenan datos. Puede ser simples o compuestos.
2. Relación: asociación entre objetos.

3. Restricción estática: propiedad estática del mundo real que no puede expresarse con los anteriores, ya que sólo se da en la base de datos; suele corresponder a valores u ocurrencias, y puede ser sobre atributos, entidades y relaciones.
4. Objeto compuesto: definidos como nuevos objetos dentro de la base de datos, tomando como punto de partida otros existentes, mediante mecanismos de agregación y asociación.
5. Generalización: se trata de relaciones de subclase entre objetos, es decir, parte de las características de diferentes entidades pueden resultar comunes entre ellas.

Por su parte, los conceptos dinámicos responden a:

1. Operación: acción básica sobre objetos o relaciones (crear, modificar, eliminar...).
2. Transacción: conjunto de operaciones que deben ejecutarse en su conjunto obligatoriamente.
3. Restricción dinámica: propiedades del mundo real que restringen la evolución en el tiempo de la base de datos.

2.4. Los sistemas de gestión de bases de datos.

Para plasmar los tres niveles en el enfoque o modelo de datos seleccionado, es necesaria una aplicación que actúe de interfaz entre el usuario, los modelos y el sistema físico. Esta es la función que desempeñan los SGBD, ya reseñados, y que pueden definirse como un paquete generalizado de software, que se ejecuta en un sistema computacional anfitrión, centralizando los accesos a los datos y actuando de interfaz entre los datos físicos y el usuario. Las principales funciones que debe cumplir un SGBD se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad. Se han señalado como componentes de un sistema ideal de gestión de bases de datos los siguientes (FROST, 1989: 90):

1. Un lenguaje de definición de esquema conceptual.
2. Un sistema de diccionario de datos.
3. Un lenguaje de especificación de paquetes de entrada/salida.
4. Un lenguaje de definición de esquemas de base de datos.
5. Una estructura simétrica de almacenamiento de datos.
6. Un módulo de transformación lógica a física.
7. Un subsistema de privacidad de propósito general.
8. Un subsistema de integridad de propósito general.
9. Un subsistema de reserva y recuperación de propósito general.
10. Un generador de programas de aplicación.
11. Un generador de programas de informes.
12. Un lenguaje de consulta de propósito general.

El SGBD incorpora como herramienta fundamental dos lenguajes, para la definición y la manipulación de los datos. El lenguaje de definición de datos (DDL, Data Definition Language) provee de los medios necesarios para definir los datos con precisión, especificando las distintas estructuras. Acorde con el modelo de arquitectura de tres niveles, habrá un lenguaje de definición de la estructura lógica global, otro para la definición de la estructura interna, y un tercero para la definición de las estructuras externas.

El lenguaje de manipulación de datos (DML, Data Manipulation/ Management Language), que es el encargado de facilitar a los usuarios el acceso y manipulación de los datos. Pueden diferenciarse en procedimentales (aquellos que requieren qué datos se necesitan y cómo obtenerlos) y no procedimentales (que datos se necesitan, sin especificar como obtenerlos), y se encargan de la recuperación de los datos almacenados, de la inserción y supresión de datos en la base de datos, y de la modificación de los existentes.

Establecidos los conceptos de bases de datos, su arquitectura y las características de las aplicaciones que soportan su gestión, es conveniente revisar los pasos o fases que sigue la ejecución de una tarea cualquiera por parte del sistema de gestión de bases de datos (MOTA, CELMA y CASAMAYOR, 1994: 13-14):

1. Petición de la aplicación del usuario.
2. Examen de la petición en el marco del esquema externo del usuario.
3. Transformación del esquema externo al lógico.
4. Transformación del esquema lógico al interno.
5. Interacción con el almacenamiento físico.
6. Envío de los datos a los buffers del SGBD.
7. Transformaciones de los datos entre el esquema lógico y el externo.
8. Transferencia de los datos necesarios al área de trabajo del usuario.

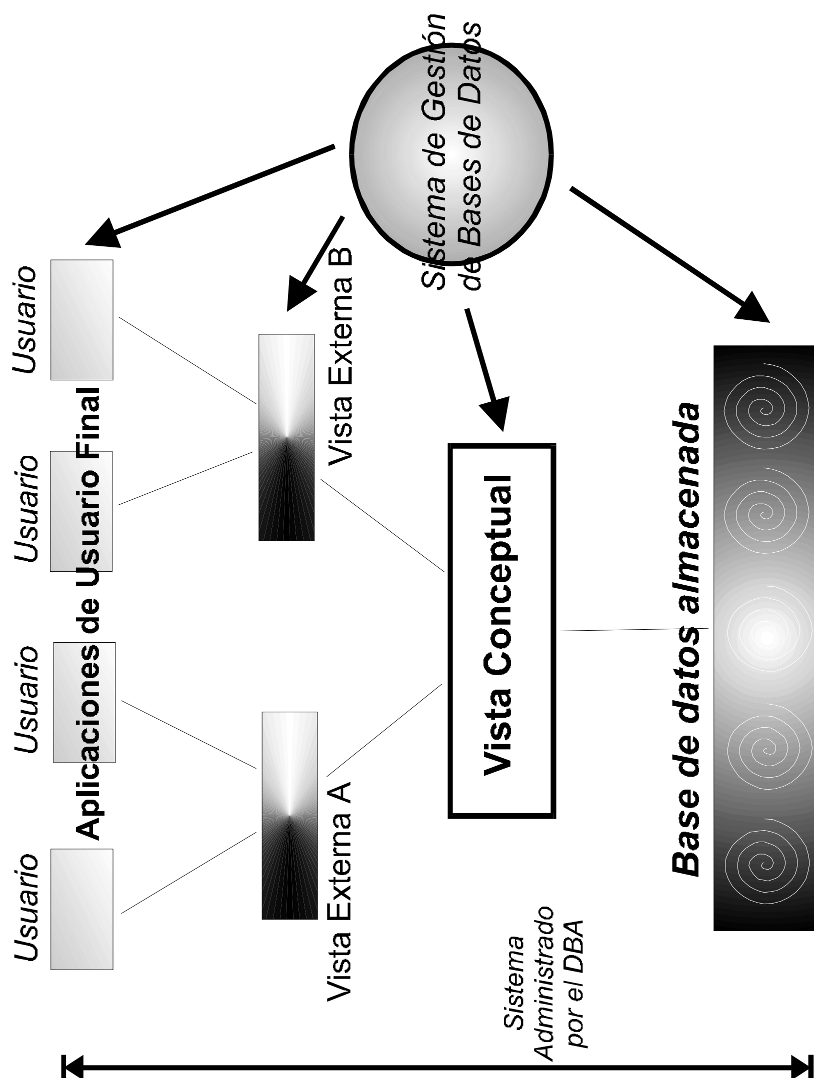


Fig.2.2. Arquitectura de un Sistema de Bases de Datos.

2.5. Los usuarios.

En consonancia con las posibles, y diferentes, vistas externas, se pueden identificar varios tipos de usuarios. En primer lugar, los usuarios finales, que hacen un uso limitado de las capacidades del sistema, normalmente referentes a introducción, manipulación y consulta de los datos. Los usuarios finales pueden ser sofisticados o especializados e ingenuos, dependiendo de su nivel de interacción con el sistema. En segundo lugar hay que citar a los programadores de base de datos, encargados de escribir aplicaciones limitadas, mediante el lenguaje

de programación facilitado por el SGBD, normalmente algún lenguaje de cuarta generación, que faciliten la ejecución de tareas por parte de los usuarios finales. Por último, el administrador de base de datos (DBA, Data Base Administrator) cumple las importantes funciones de crear y almacenar las estructuras de la base de datos, definir las estrategias de respaldo y recuperación, vincularse con los usuarios y responder a sus cambios de requerimientos, y definir los controles de autorización y los procedimientos de validación.

2.6. La creación de bases de datos.

Con los antecedentes señalados, se inicia la creación de las bases de datos. En primer lugar, y acorde con los diferentes niveles de arquitectura de bases de datos reseñados, tiene lugar la construcción del modelo y del esquema conceptual de la base de datos (REINGRUBER y GREGORY, 1994).

2.6.1.. El esquema conceptual.

El esquema conceptual puede definirse como una descripción abstracta y general de la parte o sector del universo real que el contenido de la base de datos va a representar, llamada en ocasiones "universo del discurso". En este nivel de análisis se está tratando con una descripción de la realidad, no con datos, y suele contener listas de tipos de entidades, de las relaciones existentes entre esas entidades y de las restricciones de integridad que se aplican sobre ellas. El esquema conceptual de la base de datos puede utilizarse para integrar los intereses de los diferentes usuarios, como herramienta de representación y de formación, así como para prever futuras modificaciones del sistema. En el aspecto de la representación, lo más interesante es utilizar algún tipo de especificación formal en sentido matemático, lo que facilita la consistencia y los análisis lógicos de los esquemas propuestos. Del esquema conceptual formalizado pueden derivarse diferentes subesquemas conceptuales, que representan aquellas partes del esquema conceptual de interés para un usuario o grupo de usuarios finales.

2.6.2. El esquema de la base de datos.

Una vez construido el esquema conceptual, el diseño de bases de datos obliga a realizar varias tareas previas a la construcción del esquema lógico global del sistema, también llamado esquema de bases de datos. Por el momento, basta saber que el esquema de la base de datos representa la descripción de los datos de la base de datos, mientras que el esquema conceptual representaba a la realidad. La primera de las tareas necesarias es la identificación de los datos requeridos, para obtener como resultado las partes del área de aplicación que deben representarse mediante datos, y en que forma deben presentarse éstos a los usuarios. El siguiente paso es el análisis de datos, consistente en la definición y clasificación de esos datos, su descripción, que suele presentarse en forma de diccionario de datos, como una "metabase de datos". Por último, debe hacerse la especificación de los paquetes de entrada y de salida, correspondientes con los datos que deben introducir y obtener como respuesta los usuarios, según sus necesidades. Las tres tareas habrán permitido obtener tres documentos sobre descripción del área de aplicación, definición y clasificación de los datos y

especificación de las características de los diversos paquetes, respectivamente. Tomando como punto de partida estos tres elementos, se construye la especificación de esquema de la base de datos, que responderá al contenido total de la base de datos y las características de las vías de acceso requeridas a través de estos datos. Frente al análisis de datos, que es la definición y clasificación de los datos, el esquema se encarga de la utilización de esos datos.

2.6.3. El diccionario de recursos de información (MIGUEL y PIATTINI, 1995).

La gestión efectiva de los datos involucrados en la base de datos implica necesariamente disponer de alguna herramienta que controle las características y funciones de aquéllos. Esta función es cubierta mediante el diccionario de recursos de información (DRI),que asegura la integración de toda la información contenida en el sistema. Se habla entonces de metadatos, como datos que definen y describen los datos existentes en el sistema. En un primer momento, este tipo de cuestiones eran resueltas a través de los diccionarios de datos, que reunían información sobre los datos almacenados, sus descripciones, significados, restricciones, usos, etc., y los directorios de datos, subsistemas del sistema de gestión, encargados de describir dónde y cómo se almacenaban los datos, Actualmente se aplica el concepto de diccionario de recursos de información, que engloban todo lo señalado anteriormente, dando lugar a lo que ha pasado a llamarse “metabases”.

2.6.4. El enfoque de tratamiento de los datos.

La construcción de los modelos conceptual y lógico de las bases de datos requiere la adopción de un determinado enfoque para la descripción y el tratamiento de los datos. Sin embargo, es necesario insistir en que la modelización de datos se orienta al conocimiento en profundidad de los datos que va a manejar la organización, para lograr una implantación óptima. La unión del modelo de datos con el sistema de gestión de base de datos dará como resultado la base de datos real. El modelo de datos será una representación gráfica orientada a la obtención de las estructuras de datos de forma metódica y sencilla, agrupando esos datos en entidades identificables e individualizables, y será reflejo del sistema de información en estudio.

2.7. Creación de una base de datos: enfoque E/R y transformación relacional.

2.7.1. El enfoque entidad-relación de Chen.

Por sus características, se ha seleccionado el enfoque entidad-relación propuesto por Chen (CHEN, 1976; MOTA, CELMA Y CASAMAYOR, 1994; KORTH y SILBERSCHATZ, 1993: 25-226; BATINI, CERI y NAVATHE, 1994). Este modelo toma como punto de partida considerar la existencia de entidades, que representan objetos, personas, etc, sobre las que se quiere almacenar información relevante. Las entidades con las mismas características forman un tipo de entidad. A las características necesarias para describir completamente a cada tipo de entidad se les denominará atributo. Posteriormente, las entidades y sus atributos

se representan físicamente a través de tablas (transformación en un modelo relacional) en las que los datos se almacenan en dos dimensiones. Las filas de la tabla contienen los atributos de cada una de las entidades, y las columnas el conjunto de atributos del mismo tipo de cada entidad. El grado de la tabla corresponderá al número de columnas de la tabla. En este momento estaremos trasladando el modelo semántico entidad/relación al modelo clásico relacional, se decir, la transformación entre el modelo conceptual y el lógico. El principio fundamental en este modelado, que no puede obviarse de ninguna forma, es que hechos distintos deben almacenarse en objetos distintos.

Uno de los puntos fuertes de este modelo es que prevé que las entidades puedan mantener relaciones entre ellas. En primer lugar, es necesario definir la clave de la entidad. La claves serán el atributo, o conjunto de atributos, perteneciente al mismo tipo de entidad que hacen único el acceso a esa entidad u ocurrencia de la tabla, determinando de esta forma a una única entidad. La presencia de varios atributos que pueden funcionar como clave da lugar a la existencia de claves candidatas, y por otra parte se puede hablar de claves simples (formadas por un único atributo) y claves múltiples, compuestas o concatenadas (formadas por un conjunto de atributos. No hay que obviar tampoco el concepto de clave ajena, aquel atributo de una tabla que puede funcionar como clave en otra. La ocurrencia de entidad será, en este contexto, cada uno de los posibles valores reales que puede tomar la clave de una entidad.

Las relaciones entre tablas, basadas en la conexión de éstas a través de las claves, pueden ofrecer diferentes cardinalidades, entendiendo como tal el número de ocurrencias de una entidad que se relacionan con ocurrencias de la otra entidad. Pueden identificarse tres tipos: (1,1), donde una ocurrencia se relaciona con otra; (1,m), donde una ocurrencia puede relacionarse con varias; y (m,n), donde varias ocurrencias de una entidad pueden relacionarse con varias ocurrencias de la otra entidad.

El modelo de Chen es n-ario, lo cual quiere decir que las relaciones pueden establecerse entre una, dos o más entidades. Las entidades pueden ser de dos tipos:

1. Entidad regular: aquella sobre la que se puede definir la clave primario dentro de sus propios atributos.
2. Entidad débil: aquellas que no puede utilizar sus propios atributos como clave, al estar asociada a otra entidad.

La definición del modelo conceptual con la técnica propuesta por Chen propone una secuencia de fases para la obtención del modelo:

1. Identificar las entidades dentro del sistema: para ello, debe conocerse el funcionamiento del sistema en estudio, a través de estudios de usuarios, de necesidades de información, de tipos de información, etc. Como guía puede utilizarse para la definición de las entidades objetos reales, personas, actividades del sistema, objetos abstractos, etc.

2. Determinar las claves o identificadores de entidades: señalar aquellos atributos que identifiquen inequívocamente cada ocurrencia de la entidad, y que no puedan ofrecer valores nulos.
3. Establecer las relaciones entre la entidades, describiendo el grado de las mismas: estudiar las asociaciones entre las entidades, para definir su importancia dentro del contexto del sistema, y obtener su cardinalidad.
4. Dibujar el modelo de datos: representar gráficamente el modelo obtenido.
5. Identificar y describir los atributos de cada entidad: señalar aquellas propiedades de la entidad de interés para el sistema.
6. Verificaciones: eliminación de las relaciones redundantes y que puedan ser obtenidas a través de combinar otras asociaciones.

El modelo obtenido se representa mediante una notación gráfica especializada, a través de diagramas, cuyas normas generales y variantes especializadas pueden encontrase en la bibliografía pertinente.

2.7.2. La normalización.

El modelo conceptual de datos obtenido mediante la técnica de entidad-relación será refinado y convertido en un modelo lógico relacional, utilizando la normalización, lo que ofrecerá como resultado el conjunto de tablas a implementar en la base de datos (JACKSON, 1990; MIGUEL Y PIATTINI, 1993: 425-674). Su finalidad es reducir las inconsistencias y redundancias de los datos, facilitar el mantenimiento y evitar las anomalías en las manipulaciones de datos. El objetivo será obtener un modelo lógico normalizado que represente las entidades normalizadas y las interrelaciones existentes entre ellas. Para ello, se toma como punto teórico de partida el concepto de dependencia funcional, que dice: "Un atributo B depende funcionalmente de otro atributo A, de la misma entidad si a cada valor de A le corresponde sólo un valor de B." Lo anterior se completa mediante la dependencia funcional completa y la dependencia transitiva.

El procedimiento de normalización consiste en someter a las tablas que representan entidades a un análisis formal para ver si cumplen, o no, las restricciones necesarias que aseguren evitar los problemas citados con anterioridad. A mayor nivel de normalización, mayor calidad en la organización de los datos y menor peligro para la integridad de los datos. Este procedimiento consiste en ir alcanzando formas normales

Todo el proceso se basa en que una primera relación universal plantearía enormes problemas de redundancia, consistencia e integridad de los datos, por lo que es necesario mejorar las relaciones. Estas mejoras deben dar como resultado tablas equivalentes y mejores que sus respectivas originales, y poseer siempre tres propiedades: conservación de la información (de atributos y de tuplas), conservación de dependencias y mínima redundancia de los datos. Las mejoras introducidas obligan a plantear hasta que Forma Normal es necesario llegar, es decir, a que nivel de depuración. Normalmente, es recomendable alcanzar la máxima Forma Normal, aunque luego es muy probable que restricciones existentes, de algún tipo, obliguen a retroceder a un nivel inferior de normalización, o incluso a cierto nivel de “desnormalización”.

2.8. Propuesta de un método estándar de diseño.

Con los métodos que se han expuesto, el diseño de una base de datos relacional puede seguir dos caminos. Por una parte, puede crearse tomando como punto de partida la observación del universo en estudio, dando lugar a un conjunto de esquemas de relaciones, que contengan los atributos y sus restricciones. Por otra parte, puede dividirse el diseño en dos fases, la primera de las cuales sería definir el modelo conceptual y su esquema, y la segunda transformar el esquema conceptual en un esquema relacional mediante una transformación realizada de acuerdo a unas reglas dadas.

Sin perjuicio del rigor en el diseño relacional, el diseño de una base de datos no puede limitare a la aplicación exclusiva de la teoría de la normalización. Del mismo modo que se ha visto la existencia de variadas metodologías en el ámbito de los sistemas de información, se encuentra el mismo panorama en el diseño de bases de datos, aunque aquí tampoco aparece una metodología consagrada. De esta forma, Elmasri y Navathe comparan el ciclo de diseño de los sistemas de información y de las bases de datos, y definen el problema de diseñar una base de datos como:

> "Desing the logical and physical structure of one or more databases to accommodate the information needs of the users in an organization for a defined set of applications"(ELMASRI y NAVATHE, 1989: 457)

y señalan la existencia de seis fases en el proceso de diseño de una base de datos:

Fase 1: Recopilación y análisis de requerimientos.
En esta fase se trata de conocer las expectativas del usuario sobre la base de datos. Para ello, se identifican los grupos de usuarios reales y posibles y las áreas de aplicación, se revisa la documentación existente, se analiza el entorno operativo y los requerimientos de procesado, y se realizan entrevistas y cuestionarios con los usuarios. Para todo ello existen técnicas formalizadas de especificación de requerimientos.

Fase 2: Diseño conceptual de la base de datos.
Esta fase se subdivide en otras dos. La Fase 2a corresponde al Diseño del esquema conceptual, esquema de especificación del modelo de datos a alto nivel, independiente de cualquier SGBD, que no puede utilizarse para implementar directamente la estructura de la base de datos. Para obtenerlo puede adoptarse un enfoque de esquema centralizado (en el cual se unen previamente los diferentes requerimientos a la realización del esquema), o un enfoque de integración de vistas (en el cual se unen los esquemas de cada requerimiento en uno global realizado a posteriori). La Fase 2b corresponde al diseño de transacciones, es decir, a aquellas aplicaciones que van a manipular datos

contenidos en la base de datos. Se suelen identificar mediante el estudio de las entradas y salidas de datos y su comportamiento funcional. De esta forma se identifican transacciones de recuperación, de actualización y mixtas.

Fase 3: Elección de un SGBD.
Se consideran diferentes factores técnicos, económicos y de beneficio, de servicio técnico y formación de usuarios, organizativos de rendimiento, etc. Sin embargo, resulta difícil la medida y cuantificación ponderada de los diferentes factores.

Fase 4: Transformación del modelo de datos (o fase de diseño lógico).
En esta fase se crea un esquema conceptual y los esquemas externos necesarios en el modelo de datos del SGBD seleccionado, mediante la transformación de los esquemas de modelo de datos a alto nivel obtenidos en la Fase 2a, al modelo de datos ofrecido por el SGBD.

Fase 5: Diseño de la base de datos física.
Consiste en definir las estructuras de almacenamiento y de acceso para alcanzar una rendimiento óptimo de las aplicaciones de la base de datos. Los criterios adoptados suelen ser el tiempo de respuesta, la utilización de espacio y el volumen de transacciones por minuto.

Fase 6: Implementación del sistema de base de datos.
En esta fase final se hace realidad la base de datos, mediante la creación y la compilación del esquema de bases de datos y de los ficheros de bases de datos, así como de las transacciones, a través de las aplicaciones.

La metodología expuesta, que puede servir como marco de referencia general, puede modificarse según las características del contexto en el que se diseña e implanta el sistema de bases de datos.

En el dinámico entorno de la información almacenada en las bases de datos, las recientes tendencias, derivadas en muchas ocasiones de las propias necesidades, han obligado a completar e incorporar nuevos conceptos y enfoques en el tratamiento de los datos. Por ejemplo, la existencia de relaciones complejas en el mundo real han obligado a la incorporación del modelado semántico, lo que ha dado como resultado la evolución del modelo entidad-relación extendido, con sus conceptos de superclases y subclases, y los procesos de generalización y especialización, así como la importante noción de herencia. También es necesaria la referencia ineludible al paradigma de la orientación a objetos (BERTINO y MARTINO, 1995), enfoque de tratamiento de la información que cobra cada vez mayor auge en aplicaciones comerciales, y que se configura como la opción de mayor futuro en el desarrollo de aplicaciones. La identificación de los datos y sus procesos como objetos individuales, el encapsulamiento y las propiedades de herencia son las características principales del enfoque a objetos. Por último, no puede olvidarse la creciente tendencia entre el enfoque relacional y el modelo de objetos, así como la integración de información referenciada espacialmente en modelos relacionales.

Es innegable que la gestión y la explotación subsiguiente de los registros que contienen datos, y, como consecuencia, información, depende de las herramientas existentes en el campo de la gestión de la información, por una parte, y del cuerpo teórico de la ciencia de la información, por otra. La explotación satisfactoria de esta información, de la misma forma, demanda experiencia en dos áreas de conocimiento: en las técnicas de recuperación de información y en el estudio de las necesidades de los usuarios.

3. LA RECUPERACIÓN DE INFORMACIÓN.

Lecturas recomendadas: MEADOW, 1992; GARCÍA MARCO, 1995a, 1996a; RIJSBERGEN, 1979; BLAIR, 1990; ELLIS, 1990, 1996; MACLEOD, 1991; INGWERSEN, 1992; SALTON y MCGILL, 1983; HEAPS, 1978; DESCHATELETS, 1986; BATES, 1981; BELKIN y CROFT, 1987; SARACEVIC *et alii*, 1988; CROFT, 1987; ROBERTSON, 1977; BELKIN, ODDY y BROOKS, 1982; WORMELL, 1988; FROST, 1989; GIL LEIVA y RODRÍGUEZ MUÑOZ, 1996; RIJSBERGEN y LALMAS, 1996; TURTLE y FLOOD, 1995.

3.1. Concepto de recuperación de información.

Cuando un usuario se plantea la necesidad de obtener nueva información sobre un asunto o materia de su interés, está manifestando una carencia, una situación irregular de sus estructuras mentales y cognitivas. Belkin, Oddy y Crofts han hablado de ASK, o "Anomalous State of Knowledge" (BELKIN, ODDY y CROFTS, 1982), un estado mental de incertidumbre que mueve al individuo a desarrollar una serie de acciones para salir de ese estado. Ingwersen deduce la existencia de un problema personal de espacio, "problem space" (INGWERSEN, 1992), por la diferencia entre el estado actual del conocimiento del usuario, y del estado que sería necesario para solucionar algún tipo de necesidad planteada. La respuesta a este tipo de situaciones es un conjunto de actividades que desarrolla el individuo para salir del estado anómalo, o para solucionar su problema de espacio, actividades que están íntimamente relacionadas con la adquisión de nueva información, y con el proceso comunicativo pertinente.

La recuperación de información es el conjunto de tareas mediante las cuales el usuario localiza y accede a los recursos de información que son pertinentes para la resolución del problema planteado. En estas tareas desempeñan un papel fundamental los lenguajes documentales, las técnicas de resumen, la descripción del objeto documental, etc. (CROFT, 1987) Todos estos factores ya han sido tratados exhaustivamente por la bibliografía especializada, por lo que no resulta necesario insistir sobre ellos en este lugar. Sin embargo, si es necesario recordar, como se ha señalado en la introducción, que la

informatización de estas tareas se engloba dentro del campo de la Documentación Automatizada o Informática Documental.

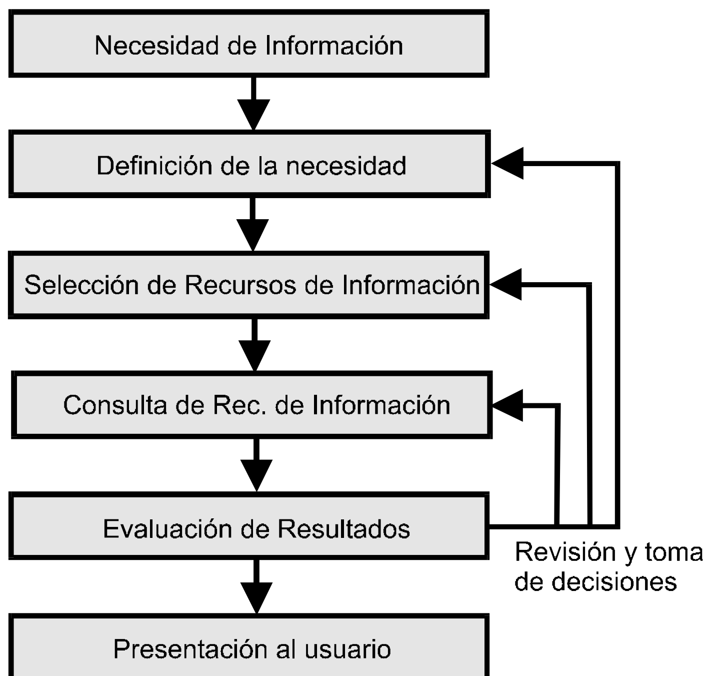


Fig.3.1. Proceso genérico de recuperación de información.

En principio, la recuperación de información engloba las acciones encaminadas a identificar, seleccionar y acceder a los recursos de información útiles al usuario, sin perjuicio de otras acepciones del concepto, en las que puede profundizarse utilizando la bibliografía correspondiente (ROBERTSON, 1977). Como puede deducirse del capítulo anterior, el objeto documental se ha organizado y representado, utilizando una serie de normas y convenciones, en un soporte informático, mediante el diseño, creación y mantenimiento de bases de datos (RIJSBERGEN, 1979). La siguiente fase lógica es la recuperación del contenido de las mismas, siguiendo unos criterios de pertinencia. En el ámbito de la documentación automatizada, la recuperación de información adoptará la forma del acceso, selección y explotación de las bases de datos, utilizando diferentes técnicas.

El planteamiento de la recuperación de información (*Information Retrieval, IR*), en su moderno concepto y discusión, hay que buscarlo en la realización de los test de Cranfield (ELLIS, 1990: 1-17), y en la bibliografía generada desde ese momento y referida a los mecanismos más adecuados para extraer, de un conjunto de documentos, aquellos que fuesen pertinentes a una necesidad informativa dada. Las propias características de las entidades del mundo real, así

como del tratamiento al que son sometidas, proveen a la representación de las mismas de un cierto grado de indefinición. Es decir, que el proceso documental, por muy alto nivel de perfección que pueda alcanzar, siempre introduce un factor de distorsión en la representación del documento. Si se considera que el acceso al documento se realiza casi por completo utilizando esta representación como intermediario, puede deducirse que los mecanismos en los que se basa la IR no son perfectos, sino que se verán influenciados por ese factor, independientemente de su validez técnica.

Los problemas de la IR resaltan todavía más cuando el usuario se sitúa en un entorno informatizado. Los niveles de definición y de relación presentes en un sistema de bases de datos relacional, por ejemplo, son más precisos y rigurosos que los existentes en un sistema de bases de datos documentales, y este rigor y control permite la recuperación de los datos necesarios en un momento dado de una forma rápida y eficaz, utilizando criterios casi inequívocos (MACLEOD, 1991). Por contra, la recuperación de información en un sistema documental, "documático", si bien a nivel técnico es rigurosa, depende en gran manera de los criterios utilizados en la representación del contenido del documento, y en los criterios utilizados en la representación ideal del documento que se utiliza en la recuperación. Influyen entonces factores ajenos al propio sistema informático, más relacionados con el intermediario humano entre el documento original y el sistema informático.

Ha sido Blair (BLAIR, 1990: 2-4) quien ha resumido las diferencias entre "data retrieval" (recuperación de datos, RD) e "information retrieval" (recuperación de información, RI), utilizando como criterios las siguientes cuestiones:

1. Según la forma de responder a la pregunta: en RD se utilizan preguntas altamente formalizadas, cuya respuesta es directamente la información deseada. En RI las preguntas resultan difíciles de trasladar a un lenguaje normalizado, y la respuesta es un conjunto de documentos que pueden contener, sólo probablemente, lo deseado, con un evidente factor de indeterminación.
2. Según la relación entre el requerimiento al sistema y la satisfacción de usuario: en RD la relación es determinística entre la pregunta y la satisfacción. En RI es probabilística, a causa del nivel de incertidumbre presente en la respuesta.
3. Según el criterio de éxito: en RD el criterio a emplear es la corrección y la exactitud, mientras que en RI el único criterio de valor es la satisfacción del usuario, basada en un criterio personal de utilidad.
4. Según la rapidez de respuesta: en RD depende del soporte físico y de la perfección del algoritmo de búsqueda y de los índices. En RI depende de las decisiones y acciones del usuario durante el proceso de interrogación.

Ha señalado este autor la importancia, en ocasiones ignorada, que tiene el factor de predicción. Predicción por parte del usuario, ya que éste debe intuir, en numerosas ocasiones, los términos que han sido utilizados para representar el contenido de los documentos, independientemente de la presencia de mecanismos de control terminológico. Este criterio de predicción es otro de los

elementos que desempeñan un papel fundamental en el complejo proceso de la recuperación de información.

3.2. El problema de los lenguajes.

Una de las cuestiones a considerar es el problema planteado por la traslación de los conceptos a diferentes tipos de lenguajes, cada uno de ellos con características propias (DESCHATELETS, 1986). En un primer momento, el usuario plantea sus necesidades utilizando el lenguaje humano, principalmente en sus variantes escrita y oral. El mensaje emitido es recibido, en numerosas ocasiones, por el documentalista, que debe asegurarse de reducir al mínimo las posibles diferencias entre lo expresado por el usuario y su propia comprensión, confrontando el significado real de los términos para el usuario final, y ofreciendo al mismo otras posibilidades: eliminación de polisemias, uso de sinonimias... Esta primera transmisión y depuración del mensaje coincide con la fase clásica de entrevista, en un proceso de búsqueda documental.

Una vez establecido claramente el contenido del primer mensaje, en un lenguaje al que podríamos llamar, con las debida precauciones "clásico", los conceptos delineados y las relaciones entre ellos deben traducirse a un lenguaje documental. Esta es la segunda transformación que se realiza sobre los conceptos, y consiste en adecuar lo expresado por el usuario, conceptos y relaciones, a los términos y relaciones propios del lenguaje documental pertinente para cada caso. Entra en juego un "segundo lenguaje". Se trata, en este momento, de traducir los términos y expresiones utilizados por el usuario, a términos y relaciones entre los términos que estén contemplados en el lenguaje documental, independientemente de su tipo, que haya sido utilizado en la creación y representación de los documentos en la base de datos que se vaya a consultar. Resulta necesario utilizar los mecanismos de control terminológico establecidos para cada recurso de información. Por supuesto, es posible utilizar directamente los términos y relaciones obtenidos del "primer lenguaje", pero puede imaginarse fácilmente que los resultados no serán los adecuados.

Sin embargo, no será ésta la última transformación "lingüística" a realizar. Una vez obtenidos y validados los términos que deban utilizarse, así como las relaciones entre ellos, las expresiones resultantes deben transformase a un "tercer lenguaje", el cual cumple la función de interrogar, de una forma consistente y comprensible para el sistema informático y sus aplicaciones, la base de datos, con la finalidad de extraer de ésta aquellos documentos que cumplan los requisitos establecidos. A este tercer lenguaje se le denomina lenguaje de interrogación de bases de datos, y es el marco dentro del cual se deben introducir las expresiones del "segundo lenguaje". En este "tercer lenguaje" los términos y conceptos expresados en los anteriores serán válidos, pero será preciso utilizar las características propias del mismo en el momento de expresar las relaciones entre los conceptos.

Los procesos de transferencia entre los tres lenguajes suponen uno de los principales problemas en la recuperación de información en bases de datos. Mientras los "terceros lenguajes" sólo ofrecen problemas de tipo técnico,

superables mediante la práctica, los procesos que rodean los "primeros y segundos lenguajes" se encuentran con un mayor nivel de azar, lo que repercute en la recuperación de información.. Para superar estos problemas se investiga en el desarrollo de nuevas técnicas que permitan superar, en lo posible, las diferencias de lógica entre los esquemas mentales de los usuarios, y los esquemas existentes en los lenguajes documentales y los lenguajes informáticos, más limitados y rigurosos. Se trataría, en este caso, de desarrollar intermediarios que facilitasen las transformaciones, de forma transparente para el usuario.

3.3. El proceso de recuperación.

Como se ha señalado anteriormente, el proceso de recuperación de información engloba numerosas tareas, de las que la consulta de recursos de información electrónica resulta ser una más de ellas (SARACEVIC *et alii*, 1988). No se van a analizar en este lugar los procesos relacionados con la definición de áreas de interés del usuario, ni con la utilización de los lenguajes documentales en la recuperación, aunque se haga referencia a estas funciones a lo largo de la exposición. Un proceso de recuperación, al que podríamos considerar "genérico" (BELKIN y CROFT, 1987) seguiría las siguientes fases:

1. Definición de las necesidades informativas del usuario.
2. Selección y ordenación de las fuentes a utilizar.
3. Traslación de las necesidades del usuario al lenguaje documental propio de la fuente a utilizar en cada caso. Es posible, además, encontrar fuentes en las que no se utilice ningún tipo de vocabulario controlado, en cuyo caso resultará necesario afinar el trabajo terminológico.
4. Traducción de la expresión de lenguaje documental al lenguaje de interrogación propio de cada sistema.
5. Ejecución de las expresiones del lenguaje de interrogación obtenidas.
6. Consulta de las respuestas obtenidas, para analizar su pertinencia o no a la cuestión planteada.
7. Replanteamiento, si procede, de las expresiones utilizadas, si los resultados obtenidos no son pertinentes.
8. Selección y obtención de los documentos que respondan a las necesidades manifestadas por el usuario.
9. Transmisión del resultado, preparado adecuadamente, al usuario.

Todas las fases son susceptibles de tratamiento informático, aunque éste queda claramente resaltado en las fases 5, 6 y 8. La perspectiva tradicional de la teledocumentación (que se expondrá en el capítulo correspondiente), ampliamente expuesta en numerosas obras de referencia, ha servido como base a la estructura de fases propuesta, aunque es necesario puntualizar que la expansión y la aparición de nuevas técnicas informáticas pueden modificar tanto el planteamiento como la ejecución de las acciones encaminadas a acceder a la información.

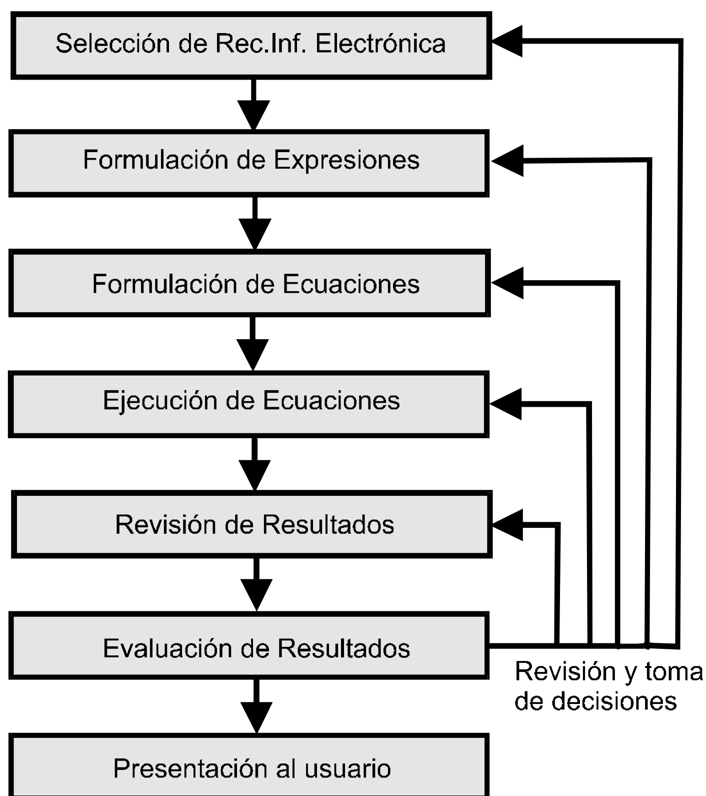


Fig.3.2. Proceso de recuperación en un entorno informático.

Resulta necesario realizar una última aclaración, referida a la utilización del término “estrategia”. La bibliografía suele utilizar esta palabra para hacer referencia a las ecuaciones utilizadas en un proceso de recuperación. Sin embargo, y desde una perspectiva tanto semántica como funcional, es más correcto utilizar el término estrategia para todo el proceso, o por lo menos para englobar las directrices generales definidas y utilizadas por el documentalista, usando el término “táctica” para las acciones más prácticas encaminadas a la obtención de los resultados. Desde este enfoque, la utilización de un conjunto de ecuaciones de consulta es más una táctica, siendo la estrategia la planificación de consulta de fuentes, los criterios de selección de las mismas, etc.

3.4. Lenguajes de interrogación y operadores.

Se puede definir a un lenguaje de interrogación como un conjunto de órdenes, operadores y estructuras que, organizados según unas normas lógicas, permiten la consulta de fuentes y recursos de información electrónica. El resultado de la combinación de estos elementos, siguiendo las normas establecidas, es una expresión, a la que se identifica con el nombre “ecuación”, capaz de interrogar el

contenido de la fuente de información. La definición mínima de un lenguaje de interrogación y de sus componentes puede encontrarse en el borrador de norma ISO 8777-1988.

Las normas lógicas que rigen un lenguaje de interrogación responden a cuestiones relacionadas con la coordinación de los elementos, es decir, con la formulación de ecuaciones. Estas normas (a modo de sintaxis) especificarán el orden de los elementos, la disposición de las estructuras, sus posibilidades combinatorias, las prioridades en la ejecución, y todo tipo de posibles funciones. Las órdenes serán aquellas palabras o abreviaturas que le indicarán al sistema las acciones a ejecutar (buscar la expresión, mostrar los documentos o registros resultantes, consultar el tesauro o los ficheros inversos, ejecutar un perfil de usuario...). Sin embargo, no todos los lenguajes de interrogación utilizan las mismas palabras como órdenes, aunque las órdenes ejecuten las mismas funciones. Existen intentos para homogeneizar la interrogación de las bases de datos, como el lenguaje CCL (Common Command Language) promovido por la Unión Europea, que aún no han alcanzado el objetivo para el que fueron desarrollados. A este panorama se une la proliferación de interfaces gráficos de usuario, que sustituyen a las órdenes y la sintaxis tradicional, dejando al usuario (si éste lo desea) sólo la labor de introducir los términos y los operadores que expresan las relaciones existentes entre ellos.

En un lenguaje de interrogación, los operadores son los encargados de expresar las relaciones que mantienen entre sí los términos que definen (más adecuado sería decir que pueden definir) las necesidades informativas del usuario. Pueden distinguirse diferentes tipos de operadores (SALTON y MCGILL, 1983), que se analizan a continuación.

3.4.1. Operadores lógicos (o booleanos).

Los operadores lógicos, llamados booleanos en honor a George Boole, precursor de la lógica simbólica y del álgebra de conjuntos, son los más utilizados en numerosos sistemas. El principio que rige la utilización de este tipo de operadores es que las relaciones entre conceptos pueden expresarse como relaciones entre conjuntos. Las ecuaciones de búsqueda pueden transformarse en ecuaciones matemáticas, que ejecutan operaciones sobre los conjuntos, lo que da como resultado otro conjunto. Los tres operadores básicos son el operador suma/unión (generalmente identificado como O/OR), el operador producto/intersección (identificado como Y/AND), y el operador resta/negación (identificado como NO/NOT). A su vez estos operadores pueden combinarse entre si, generando operaciones más complejas, como el O exclusivo (elimina la intersección), etc.

No deben obviarse los problemas que plantean los operadores booleanos, independientemente de su potencia. En primer lugar, siempre se plantean en términos de absoluto (presente/ausente), sin consideran el peso específico del término en el contexto. En segundo lugar, exigen un alto valor de precisión en los términos utilizados. Por último requieren claridad en la composición de las expresiones a buscar.

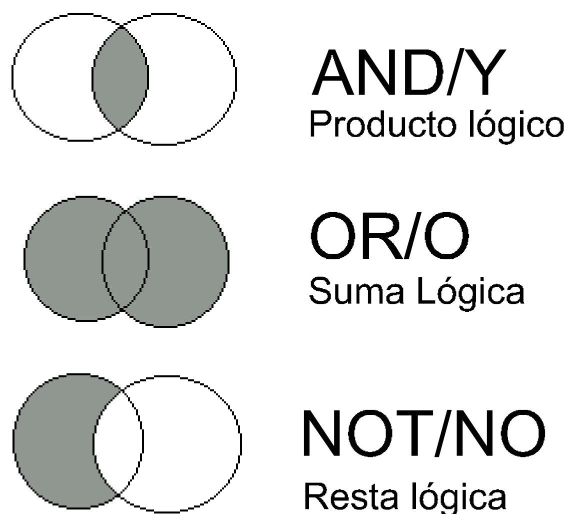


Fig. 3.3. Los tres operadores booleano básicos.

3.4.2. Operadores posicionales.

La utilización de operadores posicionales pretende superar algunas de la limitaciones que ofrecen los operadores booleanos. Toman como punto de partida la consideración del valor del término dentro del contexto, es decir, que la posición de ese término en relación con otros, o dentro del propio registro, es significativa para valorar su pertinencia a los objetivos buscados. Los operadores posicionales pueden dividirse en dos tipos:

3.4.2.1. Posicionales absolutos.

Son aquellos que permiten buscar un término en un lugar dado del documento o registro. Por regla general, son operadores de campo, es decir, permiten al usuario fijar en que campo o campos presentes en la estructura de base de datos debe aparecer el término buscado. La presencia del término en un campo dado (por ejemplo, en el campo título), puede ser una garantía de la adecuación del documento a los objetivos, en la mayor parte de las situaciones.

3.4.2.2. Posicionales relativos.

También llamados de proximidad, se trata de operadores que permiten establecer la posición de un término respecto a otro dado. Se considera que la cercanía entre los dos términos puede reflejar una íntima relación entre los conceptos reflejados por los mismos. Estos operadores permiten definir el nivel de proximidad entre los términos (mismo campo, línea, frase, número de términos significativos que los separa...).

3.4.3. Operadores de comparación.

Especifican el rango de búsqueda, fijando unos límites para la misma. Estos límites pueden ser tanto numéricos como alfabéticos, correspondiendo los operadores a formas del tipo “mayor que”, “menor o igual que”. Se utilizan principalmente en documentos que pueden contener datos numéricos.

3.4.4. Operadores de truncamiento.

Pueden darse situaciones en las cuales sea necesario utilizar no un término simple, sino también sus derivados, fijados por prefijación o sufijación, mínimas variantes léxicas, etc. Para facilitar la búsqueda de este tipo se han introducido operadores de truncamiento, a los que también se llama máscaras. Se trata de operadores (normalmente símbolos como *, $), cuya presencia puede sustituir a un carácter o a un conjunto de caracteres, situados a la izquierda, dentro o a la derecha del término en cuestión.

En los actuales sistemas de recuperación de información es posible encontrar todos estos tipos de operadores, que pueden combinarse entre sí, permitiendo crear ecuaciones complejas que reflejan con bastante precisión los conceptos y sus relaciones. La combinación de los operadores debe respetar un conjunto de reglas, básicas en todos los sistemas, que establecen las prioridades y formas de ejecución de ecuaciones complejas, cuando éstas combinan más de dos conceptos. En primer lugar, los sistemas tienden a resolver, a ejecutar en primer lugar, aquellas expresiones que se relacionan utilizando el operador más restrictivo o prioritario. Por ejemplo, un operador posicional absoluto posee un nivel de restricción (una prioridad) mayor que un operador booleano, lo que significa que el sistema ejecutará antes la expresión cuyo operador es el posicional absoluto, combinando posteriormente el resultado con el operador booleano y su término relacionado. Sin embargo, pueden darse expresiones en las cuales sea necesario variar estas prioridades, y ordenar al sistema que ejecute en primer lugar expresiones con operadores de menor nivel de restricción, relacionando luego su resultado con términos a través de operadores más restrictivos. Para estas situaciones, se utilizan paréntesis, los cuales engloban a las expresiones que deben ejecutarse en primer lugar, independientemente de las prioridades fijadas por el sistema. La utilización de expresiones entre paréntesis hace posible, por ejemplo, que el resultado de una expresión con un operador booleano pueda ser combinada con un operador posicional absoluto. Además, los paréntesis pueden anidarse, resolviéndose las ecuaciones planteadas desde dentro hacia fuera, de la misma forma que las igualdades y polinomios matemáticos.

3.5. Estrategia de la interrogación.

Los lenguajes, sus órdenes y operadores son utilizados dentro del proceso de recuperación de información, la cual se encuentra almacenada en un repositorio, que suele ofrecer la forma de base de datos. La base de datos es consultada mediante la ejecución de búsquedas, expresiones que reúnen los elementos citados con anterioridad, y cuya resolución da como resultado aquellos

elementos que responden a la lógica expresada en la búsqueda. El término "estrategia", en lo que se refiere a la consulta de bases de datos, ha servido para identificar diferentes enfoques y conceptos, que engloban desde la visión general del proceso hasta la formulación de ecuaciones individuales

La estrategia debe ser un plan ideal de interrogación de la base de datos que incluya el objetivo de la búsqueda, el plan general y el plan específico de operación. El objetivo de la búsqueda se obtiene identificando que tipo de información se necesita y sus características. Una vez definido el objetivo, debe establecerse un plan general de operación, que incluya una selección de la base o bases de datos a consultar, las primeras aproximaciones a los términos a utilizar en las ecuaciones, así como las posibles relaciones lógicas. El plan específico de operación se pone en marcha una vez obtenidos los resultados del anterior, y debe formular ecuaciones y utilizar términos con el mayor grado de precisión, establecer una secuencia lógica con todo ello, y redefinirlo si es preciso. Independientemente de ambos planes, resulta necesario conocer con anterioridad la respuesta a varias cuestiones que afectan a la interrogación de la base de datos, tales como el contenido y alcance de la base de datos, coste de consulta, lenguaje y operadores a utilizar durante las consultas, límites preestablecidos (por el usuario o el sistema)... todas ellas afectan y modifican el enfoque del interrogador.

3.5.1. Tipos de estrategia.

En el momento actual, parece más adecuado utilizar el término para identificar el plan general de búsqueda. No existe una única ni perfecta aproximación a las estrategias de interrogación de bases de datos. En la mayor parte de las ocasiones depende de la experiencia del usuario y de la calidad del contenido de los registros existentes en la base de datos, especialmente en lo que corresponde a su control terminológico. La estrategia depende, en gran manera, de la formación, intuición y experiencia del usuario. Tomando en consideración la intención del interrogador, la bibliografía señala que pueden varios tipos principales de búsqueda, que pueden clasificarse en dos grandes grupos, sin perjuicio de que puedan darse situaciones en las que se combinen (MEADOW, 1992: 243-251; BATES, 1981; TURTLE y FLOOD, 1995):

1. Categorización por objetivo:
- Búsqueda de elemento conocido: se trata de búsquedas en las cuales el interrogador sabe cual será la respuesta, cuyo contenido, por ejemplo, utiliza para completar una referencia bibliográfica, o utiliza un número de inventario para comprobar el contenido de un registro dado.
- Búsqueda de información específica: el interrogador busca una información específica dada, generalmente sobre un tema concreto y limitado, como trabajos publicados en un año o por un autor.
- Búsqueda de información general: intenta buscar la información sobre una materia o asunto, de forma general, que obtenga una visión global del estado de la misma.

- Exploración de la base de datos: se trata de conocer que tipos de información y/o documentos se encuentran almacenados en la base de datos, a qué pueden responder y cómo pueden utilizarse.
2. Categorización por plan de operación:
- Búsqueda directa: se trata de una aproximación expeditiva, en la que se intenta resolver el problema con la formulación de una única consulta. Como puede deducirse, resulta difícil obtener buenos resultados con la misma.
- Búsqueda “breve”: es una evolución de la anterior, en la que se trata de recuperar unos ítems significativos entre un gran número obtenido tras una sola ecuación.
- Ampliación: comienza con ecuaciones muy restrictivas, que ofrezcan documentos pertinentes. Tras analizar la respuesta, el usuario puede ampliar o expandir las ecuaciones de búsqueda hasta recuperar toda la información existente. Puede ofrecer problemas si la ecuación inicial no es adecuada.
- Restricción: opuesta a la anterior, formula ecuaciones que ofrecen resultados muy amplios, para posteriormente utilizar ecuaciones más restrictivas, hasta delimitar los documentos pertinentes.
- Construcción de bloques: intenta establecer bloques de información que respondan a los componentes de su lógica, para combinarlos entre sí posteriormente, hasta encontrar una combinación que responda a las necesidades planteadas.

3.6. La exploración como mecanismo de recuperación.

Las limitaciones inherentes al proceso de recuperación mediante ecuaciones han conducido a experimentar otras aproximaciones. Una de las más utilizadas es aquella que utiliza la exploración, es decir, el acceso a los documentos mediante técnicas de visualización de parte de su contenido que puede ser relevante, y la posterior asociación con otros documentos de perfil similar. Para Doyle (ELLIS, 1990: 22-23), esta capacidad de exploración debería ser fundamental en los sistemas. El usuario accede a un listado o enumeración de elementos descriptivos, y mediante un proceso de selección de elementos, va centrando el objetivo de su búsqueda. Los criterios utilizados por el usuario se basan en la deducción y la asociación de conceptos (aproximación ésta similar a la que utiliza un sistema hipertextual -ver *infra* el capítulo pertinente-) frente a la lógica de conjuntos que se plantea en un sistema de ecuaciones. Este tipo de representación es más adecuada para reflejar la polirepresentación que un concepto puede tener para un usuario individual. En cambio, la utilización de la exploración suele realizarse en entornos en los cuales el usuario no posee una idea clara de cual debería ser la mejor táctica para aproximarse a la información que precisa. Por lo tanto, la cuestión clave a considerar en un sistema de exploración es combinar las ideas y esquemas del usuario con el esquema de organización de la información que ofrece el sistema (INGWERSEN, 1992: 135-140). Esta es la aproximación que pretenden desarrollar los enfoques cognitivos, poniendo su énfasis en el intermediario que debe existir entre el modelo del usuario y el modelo del sistema.

3.7. Revisión y análisis de resultados.

El resultado de la ejecución de una ecuación de búsqueda es un conjunto de documentos que cumplen las condiciones expresadas en la ecuación. Se trata, a su vez, de un subconjunto del conjunto total de documentos existentes en el recurso o fuente de información consultado. Sin embargo, puede darse el caso de que la respuesta sea un número excesivamente elevado de documentos, o un número mínimo. Por otra parte, los documentos resultantes responden a la lógica y a las condiciones expresadas en la ecuación de búsqueda, lo cual no supone, como ya se ha señalado, que sean pertinentes a las necesidades del usuario. En realidad, es posible ejecutar ecuaciones perfectas, desde un punto de vista funcional (operadores, términos...), sin que los documentos resultantes reúnan las características que los harían deseables para el usuario.

Para superar esta posible distorsión en los resultados es necesario valorar y evaluar la respuesta a las ecuaciones planteadas. La primera modificación a realizar en la formulación de las ecuaciones afecta al número de respuestas obtenidas. En el caso de un excesivo número, se utilizan técnicas de restricción, mediante la introducción de términos más específicos, desechar términos generalistas, o limitar los truncamientos. En el caso de un número muy reducido, las acciones a tomar son las contrarias: utilizar términos más generales, incluyendo derivados y relacionados, limitar los operadores más restrictivos, añadir truncamientos, etc. Si se da la situación de ecuaciones correctas funcionalmente, pero sin respuesta adecuada, sería necesario replantear el proceso de recuperación, especialmente en la utilización de los lenguajes documentales y en la selección de fuentes.

Los resultados de una búsqueda se pueden valorar cuantitativamente utilizando dos parámetros, que son la llamada y la precisión (GARCÍA MARCO, 1996a: 63). Estos parámetros combinan el número de documentos pertinentes y no pertinentes, y recuperados y no recuperados, según las siguientes fórmulas (HEAPS, 1978: 28):

	Pertinentes	No pertinentes
Extraídos	a	b
No extraídos	c	d

La tasa de llamada responde a la fórmula a/(a+c), y debería situarse entre el 0,6 y el 0,8. La tasa de precisión responde a la fórmula a/(a+b), y se sitúa entre 0,2 y 0,8. El principal problema para calcular la tasa de llamada es conocer la variable c, casi imposible de ajustar en situaciones normales de recuperación de información, por lo que se suelen utilizar técnicas estadísticas y de muestreo para obtener valores fiables.

3.8. Recuperación de información y sistemas expertos.

Los sistemas expertos, con su capacidad para combinar información y reglas de actuación, han sido vistos como una de las posibles soluciones al

tratamiento y recuperación de información, no sólo documental. La década de 1980 fue prolija en investigación y publicaciones sobre experimentos de este orden, interés que continua en la presente década. En resumen, un sistema experto es una aplicación capaz de realizar las tareas propias de un experto humano en un área restringida. Se compone de una base de datos, de una base de reglas y de un motor de inferencia (FROST, 1989). La base de datos almacena el conjunto de datos o documentos sobre los que se desea ejecutar una serie de acciones. La base de reglas contiene un compendio de reglas lógicas que el sistema debe utilizar para desarrollar razonamientos, así como las normas que permiten combinar las reglas, por último, el motor de inferencia es el encargado de ejecutar las órdenes del usuario, utilizando como criterios las reglas, y como material de partida el contenido de la base de datos, hasta alcanzar una conclusión simulando el razonamiento que seguiría el experto humano. El desarrollo posterior de estos sistemas ha traído un conjunto de nuevas aplicaciones a las que se identifica como Sistemas Basados en el Conocimiento (SBC), que incorporan técnicas más sofisticadas como la lógica difusa, razonamiento basado en modelos, etc., y lo que puede ser más interesante para el especialista en información, medios de recuperación de información deductiva (FROST, 1989: 6).

Lo que diferencia a estos sistemas de un sistema tradicional de recuperación de información es que estos últimos sólo son capaces de recuperar lo que existe explícitamente, mientras que un sistema experto debe ser capaz de generar información no explícita razonando con los elementos que se le dan (WORMELL, 1988). Pero la capacidad de los SE y de los SBC en el ámbito de la recuperación de la información no se limita a la recuperación. Pueden utilizarse en ayudas al usuario, en selección de recursos de información, en filtrado de respuestas... (ALBERICCO y MICCO, 1990) Un SE/SBC puede actuar como un intermediario inteligente que guía y apoya el trabajo del usuario final. Para desempeñar de forma adecuada esta tarea, los enfoques centrados en la creación de modelos de las estructuras cognitivas del usuario son los más prometedores (INGWERSEN, 1987).

Página en blanco

4. SISTEMAS INFORMÁTICOS DE TRATAMIENTO Y RECUPERACIÓN DE INFORMACIÓN DOCUMENTAL.

Lecturas recomendadas: SOERGEL, 1985; TRAMULLAS y CUBILLO, 1995; TRAMULLAS, 1996b; SALTON, 1989; MEADOW, 1992; SAFFADY, 1989; WILLET, 1988; WEITZMAN y MATTHEW, 1995; ASHFORD y WILLET, 1988; CODINA, 1990; CODINA y ABADAL, 1992; DEOGUN y RAGHAVAN 1988; GREEN, 1993; LASSOURY, 1994; PRAX, 1994; REID, 1990; LITC, 1994; LANCASTER y WARNER, 1993; DEWIRE, 1994; CONLON y CONLON, 1996; TAGUE SUTCLIFFE, 1996; EWERS, 1994; ELLIS, 1996.

4.1. La gestión informática de documentos.

Son corrientes, en la actualidad, numerosas denominaciones como Sistemas de Gestión Documental, Sistemas de Archivo Electrónico, Sistemas de Recuperación de Información, Sistemas de Gestión de Bases de Datos Documentales... todas ellas se refieren a aplicaciones que, en mayor o menor medida, incorporan las funciones típicas de las Ciencias de la Documentación a un entorno informático (WILLET, 1988). Sin embargo, es corriente encontrar todo tipo de variaciones, tanto de estilo como de rango, entre la amplia panoplia de aplicaciones que se encuadran tras estos términos. Una breve discriminación terminológica de todos los términos señalados indica que el más adecuado para cubrir los contenidos de este capítulo, analizando las cuestiones planteadas por Reid y por Deogun y Raghavan (REID, 1990; DEOGUN y RAGHAVAN, 1988) es el concepto de "sistema de recuperación de información y de gestión de bases de datos documentales", ya que el resto resultan más generales, o todavía delimitados de forma somera, mientras que el objeto de estudio, en este caso, se va a centrar en las aplicaciones capaces de crear, mantener e interrogar una base de datos documental, algo que no todas las aplicaciones intituladas como se ha explicado son capaces de realizar.

Los sistemas de recuperación de información (*Information Retrieval Systems*) son aquellos que ofrecen al usuario mecanismos para acceder a fuentes de información en soporte informático, y recuperar y extraer de las mismas aquellos documentos cuyo contenido sea capaz de responder a una cuestión

planteada por el usuario (SOERGEL, 1985). Se trata de terminología anglosajona, que en origen también se utilizaba para englobar sistemas capaces de crear y gestionar fuentes de información, aunque en la actualidad es preferible usar el término para aquellos sistemas que se limitan a la recuperación, como por ejemplo en la consulta de bases de datos en CD-ROM. Otro tipo de sistema que se está extendiendo progresivamente, sobre todo en entornos empresariales y de organismos públicos, es el identificado como sistema de gestión documental, cuya finalidad es establecer un control de toda la documentación recibida y generada por una organización, independientemente de recibir un tratamiento documental completo o no. Este último término resulta excesivamente amplio, en muchos casos, dadas las prestaciones ofrecidas por los sistemas.

Los sistemas de tratamiento y recuperación de información documental (STRID) manipulan documentos, en su más amplia acepción, frente a los sistemas de bases de datos (SGBD), que manipulan registros (TRAMULLAS y CUBILLO, 1995). Limitándose al contexto de la Documentación, podría decirse que los STRID están trabajando con construcciones cognitivas que representan creaciones humanas diferentes a las que representan los registros presentes en los SGBD, más estáticos y sujetos a normas fijas. Esto no es óbice para poder utilizar un SGBD para una finalidad documental, ya que los fundamentos teóricos de las bases de datos, como se ha visto anteriormente, son iguales para ambos, pero si debe considerarse que las técnicas de manipulación de unos y otros serán más o menos adecuadas para tal fin. Sin embargo, hay que concluir que el desarrollo de la tecnología de bases de datos es continuo, y que su propia dinámica está favoreciendo, cada vez en mayor medida, la aparición de aplicaciones mixtas, y la adopción de técnicas de unas en otros, por lo que el panorama de los STRID, tan claro hace pocos años, está difuminando sus límites, al igual que sucede en otros ámbitos de la informática.

4. 2. Tipos de sistemas de tratamiento y recuperación de información documental.

La recuperación de documentos a través de ecuaciones de búsqueda es posible gracias a que existen unas aplicaciones que no sólo las ejecutan, sino que son capaces de gestionar, en el sentido más amplio, las bases de datos documentales. El proceso de especialización que dio lugar a la informática documental ha producido, a su vez, aplicaciones especializadas en diferentes aspectos* , que en muchas ocasiones entrecruzan sus características (CODINA, 1990; CODINA Y ABADAL, 1992):

- Sistemas de gestión de bases de datos documentales: son sistemas que incorporan todas las características de los SGBD tradicionales, incluyendo la creación y mantenimiento de bases de datos

* La dinámica de los productos informáticos hace difícil analizar detenidamente las aplicaciones existentes en el mercado. A tal fin, resulta recomendable consultar regularmente publicaciones especializadas de ámbito general, en las que aparecen reseñas e informes de este tipo de aplicaciones, como por ejemplo la revista *Byte*.

documentales, usuarios, controles de seguridad, e incluso lenguajes propios de programación. Debe citarse BRS/Search, de BRS Information Technologies (uno de los más completos), Inmagic, CDS-Isis y su interfaz WinIsis, Texto...

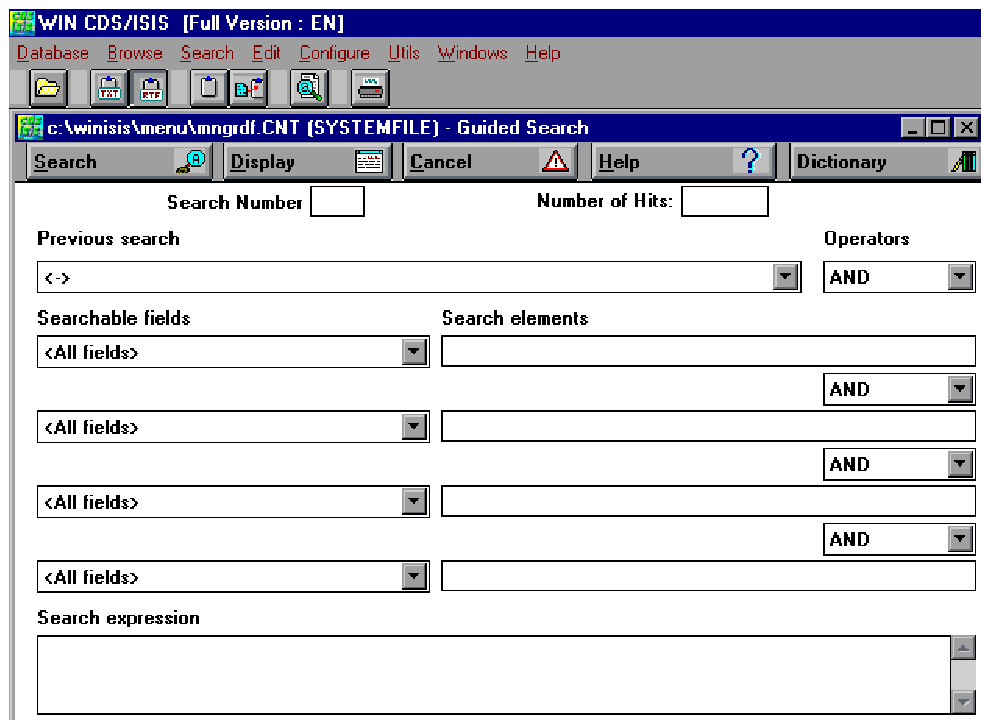


Fig.4.1. Interfaz de WinIsis

- Sistemas de indización: son aquellos que crean ficheros de índice, los cuales contienen los términos existentes en diferentes tipos de ficheros, como por ejemplo ASCII o cualquier procesador de textos, y permiten una limitada recuperación a través de aquéllos, que suele utilizar operadores booleanos. Uno de los más conocidos es ZyIndex.

- Sistemas de exploración o escáneres: se trata de aplicaciones que, sin necesidad de crear ficheros inversos ni diccionarios, son capaces de acceder a ficheros con diferentes formatos, y buscar dentro de los mismos las cadenas de caracteres que respondan a lo expresado en la ecuación de búsqueda. Pueden encontrarse aplicaciones que combinen la exploración con la indexación, como dtSearch.

- Sistemas de gestión bibliográfica: sistema especializado para la gestión y mantenimiento de bibliografías especializadas (TRAMULLAS, 1996b), es

una aplicación específica de los sistemas de gestión de bases documentales, que permiten no sólo el almacenamiento y la recuperación de referencias bibliográficas, sino también la exportación de estas referencias en diferentes formatos de cita bibliográfica a diferentes procesadores de textos, sistemas de gestión de bases de datos, etc. Como ejemplo se pueden citar Reference Manager, ProCite, EndNote...

- Sistemas de recuperación de información: son aplicaciones que se encargan exclusivamente de recuperar información de bases de datos documentales no modificables. Ponen a disposición del usuario potentes herramientas de búsqueda y de apoyo a la búsqueda, pero su funcionalidad queda reducida a la consulta y exportación de documentos. Las bases de datos que vienen en soporte CD-ROM utilizan este tipo de sistemas, de los que debe citarse SPIRS (Silver Platter Information Retrieval System) y WinSpirs (versión para Windows de Spirs), así como los proveedores de bases de datos en línea, como DIALOG.

- Sistemas hipertextuales: en su origen, los hipertextos e hipermedias son una forma de organizar, acceder y explorar documentos de diferentes tipos, que posteriormente se ha popularizado como motor y parte de tutoriales y presentaciones. Actualmente estos sistemas están volviendo a ser considerados como una forma válida y muy avanzada de gestionar documentación (véase *infra* el capítulo correspondiente).

- Sistemas de Gestión Documental o de Gestión Electrónica de Documentos (GED): se trata de sistemas que pretenden ofrecer una solución integral para la documentación, especialmente administrativa y de gestión, que se utiliza en una organización dada (PRAX, 1994; LASSOURY, 1994). Incorporan funciones clásicas de gestión de bases de datos, y utilizan esquemas de obtención de una copia del documento original mediante escáner, almacenamiento óptico o magneto-óptico, y un nivel básico de descripción textual del documento y de su contenido.

- Sistemas ó Gestores de información personal (Personal Information Systems/Managers): son aquellos que integran, en un único entorno, todos los documentos, ficheros y relaciones entre ellos que son de interés para el trabajo de un usuario. Numerosos sistemas integrados de informatización ofrecen a sus usuarios un acceso homogéneo a los diferentes tipos de documentos y ficheros que manejan en su trabajo diario.

- Sistemas compuestos: se denomina así a aquellos que dan soporte a todas las tareas que se realizan en una unidad informativa, sea ésta un archivo, biblioteca o centro de documentación. Esto significa que cubren tanto la cadena documental como la gestión administrativa. Sirvan como ejemplo las aplicaciones de automatización de bibliotecas, como Absys o Libertas, o las aplicaciones de automatización de archivos, como la desarrollada para el Archivo de Indias de Sevilla. Normalmente, integran un motor documental, encargado de gestionar las bases de datos

documentales que cubren los catálogos, y un motor relacional, que cubre las tareas administrativas (véase *infra* el capítulo dedicado a una aplicación concreta como es la informatización de bibliotecas)..

4.3. Especificidad de los STRID.

Evidentemente, algún tipo de características debe diferenciar a los sistemas de tratamiento y recuperación de información, como tales, de los sistemas de gestión de bases de datos, en el sentido más tradicional del término, vistos en el capítulo anterior. Como Codina y Abadal han señalado, en primer lugar los SGBD se centrarían en almacenar información correspondiente a procesos de gestión, fácilmente predecibles (aunque sería necesario matizar esta afirmación). Por oposición, los STRID se harían cargo de datos e informaciones poco estructuradas, con un fuerte componente textual, y de más difícil predicción, a causa de su contenido (MEADOW, 1992; ASHFORD y WILLET, 1988; SAFFADY, 1989). Las tres características que permiten identificar a un STRID son:

1. Utilización de la técnica del fichero inverso.
2. Aplicación de la lógica booleana como método de recuperación de información.
3. Manipulación de registros de extensión variable o indefinida.

4.3.1. La utilización del fichero inverso.

En primer lugar, la utilización de la técnica del fichero inverso (o fichero invertido) es un elemento clásico de los STRID. Dada la gran cantidad de información textual contenida en las bases de datos documentales, los procedimientos clásicos de búsqueda secuencial o de ficheros indexados no son capaces de responder de manera adecuada a los requerimientos de velocidad y exactitud en la respuesta necesarios para satisfacer al usuario. Por esta razón, los STRID utilizan una especialización de los ficheros indexados, creando un nuevo tipo de fichero, al que se conoce como "fichero inverso."

El fichero inverso contiene, ordenadas alfabéticamente, todas y cada una de las palabras o términos con significado semántico, contenidos en los documentos presentes en la base de datos documental. Valdría la imagen de una larga lista de palabras ordenadas, acompañada por una indicación del documento en el que aparece cada una de ellas. En el momento de realizar una búsqueda o consulta, el STRID no procedería a leer todos y cada uno de los documentos: simplemente, buscaría en el fichero inverso correspondiente todas y cada una de las ocurrencias de los términos buscados, ofreciendo como resultado el número de documentos en el cual aparecen. Dependiendo del sistema utilizado, acto seguido pueden mostrar, o no, los documentos pertinentes. En la mayor parte de las situaciones, el sistema queda a la espera de nuevas órdenes del usuario, sean de ejecutar nuevas búsquedas, sean de visualización de los documentos.

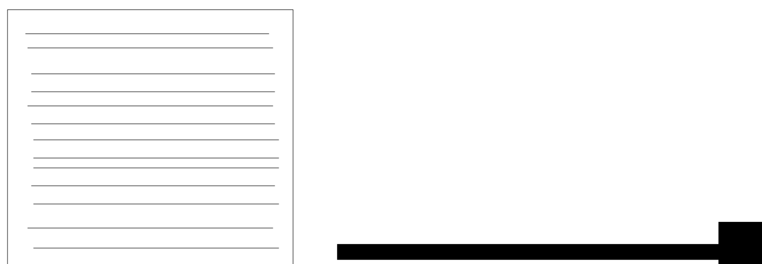


Fig.4.2. Tratamiento del documento en un STRID

En la técnica del fichero inverso se ha introducido progresivamente mejoras y especializaciones. Así, los ficheros inversos no suelen limitar su contenido al término y el documento o documentos en los que aparece. Se han incorporado también los datos referidos al campo y parágrafo, dentro del documento en el que aparece, así como la línea y la posición absoluta dentro de la línea, en el modo "documento 3, campo título, línea 2, palabra 4".

4.3.2. La lógica booleana.

El álgebra booleana es el resultado de la obra de Boole, que estableció las bases de las operaciones a través de operadores lógicos. La combinación de los términos mediante operadores lógicos crea una ecuación, la cual puede ser transformada por medios informáticos, como se ha visto en un capítulo anterior. El STRID selecciona y compara, mediante la aplicación de la teoría de conjuntos, los resultados de la consulta al fichero inverso, y selecciona, de acuerdo a los operadores establecidos por el usuario, aquellos que responden a la lógica de la ecuación de búsqueda.

4.3.3. Manipulación de registros de estructura y extensión variable.

Por último, la capacidad para representar y manipular registros de estructura y longitud variable es otra de las características propias de los STRID. Como es previsible, no todos los documentos poseen la misma estructura, ni ofrecen la misma longitud. En los SGBD tradicionales se ha considerado necesario que la estructura y contenido de todos los registros fuese similar. Sin embargo, las características propias de los documentos obligan a que el sistema de bases de

datos documentales sea capaz de aceptar, almacenar y manipular registros de forma flexible.

4.4. Tratamiento del texto.

El factor crucial que determina la calidad y el éxito de un STRID es su capacidad para manipular la información textual, tanto en lo que se refiere a su adquisición y tratamiento, con vista a su posterior aprovechamiento, como en lo que respecta a la recuperación de la información contenida en el mismo (MEADOW, 1992; SALTON, 1989; LANCASTER, 1986). Mayor importancia alcanza este factor cuando el usuario debe aceptar que la presencia de un término en un documento no asegura la pertinencia del mismo al objetivo de la búsqueda (por ejemplo, la frase “este documento no versa sobre bibliotecas”). En gran parte de las ocasiones, el texto que se introduce en la base de datos no es sometido a un riguroso control sintáctico y terminológico. Esto obliga a disponer de sistemas que no sólo se centren en el término, sino que sean capaces de reconocer patrones de texto, para superar estas limitaciones.

La primera acción ejecutada por el sistema cuando recibe un nuevo documento o registro es su indización o indexación. El sistema trata de identificar individualmente a cada uno de los términos presentes en el documento que poseen significado propio, eliminando aquellos cuya frecuencia suele ser tan alta que no son significativos, y que corresponden a lo que se engloba bajo el término “palabras vacías” (artículos, conjunciones...). Para facilitar el trabajo, los sistemas utilizan directamente una lista preexistente con esos términos, lo que agiliza la tarea. Sin embargo, la indización no permite calibrar el peso específico de un término en un documento, lo que podrá ser llevado a cabo por el sistema con posterioridad, tomando como punto de inicio el propio índice.

La información obtenida en el proceso de indización, que es introducida en el fichero inverso como se ha señalado anteriormente, debe ser lo más completa posible. No debe limitarse a la presencia o ausencia del término, sino que debe ofrecer datos sobre la posición absoluta del término en el contexto del documento. Esta información va a ser la que permita efectuar búsquedas sobre el texto que utilicen criterios de presencia, posición absoluta y proximidad. Sin embargo, es necesario disponer de otras técnicas que completen los mecanismos de recuperación de información, y que se engloban en técnicas de asociación de términos y técnicas de asociación de documentos.

Las técnicas de asociación de palabras o términos intentan buscar y ofrecer al usuario términos relacionados con los que éste expresa en su consulta. Los más conocidos son los diccionarios y los tesauros y mini-tesauros. Los diccionarios ofrecen significados, sinónimos y antónimos del término deseado, mientras los tesauros ofrecen listas de términos relacionados, según varios criterios, con el término en cuestión. Junto a éstos deben citarse las estadísticas de coocurrencia, que establecen cuando unos términos aparecen asociados a otros en los documentos de la base de datos, y los truncamientos de raíz, que permiten asociar una raíz semántica con todas las variantes de la misma

presentes en la base de datos.

Las técnicas de asociación de registros o documentos facilitan recuperar un registro o documento, y ejecutar una serie de cálculos estadísticos para ofrecer al usuario aquellos cuyo contenido es similar matemáticamente, a través del análisis de las estadísticas de coocurrencia. Deben citarse las medidas de similaridad, que se aplica a dos documentos analizando la frecuencia de las palabras en ambos; el clustering, utiliza la técnica de agrupación de registros similares en grupos; el análisis de marcas, que transforma los términos en valores numéricos, formando una representación del documento y agrupa los similares; y los métodos discriminantes, que utilizan un conjunto limitado de términos para incluir o excluir los documentos de los grupos.

4.5. Las funciones técnicas de un STRID.

Los Sistemas de Tratamiento y Recuperación de Información Documentales deben ofrecer a sus usuarios, como especializaciones de los SGBD que son, las mismas funciones técnicas de creación, mantenimiento y recuperación de la información que éstos (TRAMULLAS y CUBILLO, 1995; WEITZMAN y MATTHEW, 1995). Esto supone tanto la disponibilidad de funciones para el usuario final, de forma que tenga a su alcance todos los mecanismos necesarios para acceder a los documentos, como la existencia de herramientas y utilidades que aseguren la adecuada administración y mantenimiento del sistema, en sus diferentes niveles, por parte del DBA. Ambos deben ser complementados por nuevas capacidades, más adecuadas a un tratamiento documental e informativo de la base de datos (CONLON y CONLON, 1996). En esta situación, un STRID debe ser capaz de ofrecer las siguientes funciones:

1.- Creación de bases de datos documentales, con una estructura de campos (también llamados parágrafos o párrafos) fija. Los diferentes campos, para lograr un correcto tratamiento de los datos, deben ser de longitud variable y, a ser posible, permitir la creación de subcampos o subparágrafos. Todos los sistemas deben respetar la arquitectura de bases de datos explicada en el capítulo correspondiente.

2.- Introducción de datos, tanto de manera interactiva, como de ficheros preexistentes, con posibilidad de conversión de formatos de bases de datos externas. Las funciones de introducción de datos deben permitir controles de entrada, validación, etc, de cara a asegurar, en lo posible, la calidad de los datos introducidos.

3.- Recuperación de la información contenida en los documentos de la base de datos a través de cualquier término existente en ella, mediante la formulación de ecuaciones de búsqueda que permitan combinar los términos según diferentes criterios. Los sistemas ofrecen la posibilidad de ejecutar las consultas sobre una o varias bases de datos simultáneamente. Los documentos resultantes se agrupan en sets o conjuntos, susceptibles de combinación posterior.

4.- Salida de la información, mediante edición en pantalla, impresión y redirección a ficheros de los documentos de interés para el usuario. Las órdenes de salida de información deben ofrecer la posibilidad de enviar ésta a diferentes destinos, así como los formatos de presentación de los datos a utilizar (tamaño, campos...). Deben incluirse aquí las capacidades para ordenar, según diferentes criterios, los documentos resultantes. Otra función a considerar es la posibilidad de crear nuevas bases de datos, tomando como base los documentos recuperados en una búsqueda previa.

5.- Análisis de la respuesta, utilizando herramientas de análisis de frecuencias de los términos o de coocurrencias.

6.- Creación de ficheros que definan el perfil de búsqueda de los usuarios, así como de las ecuaciones que ejecuten. A los ficheros que contienen ecuaciones de búsqueda previamente grabadas, y que pueden ejecutarse en cualquier momento, se les llama normalmente "macros". Estas macros son ficheros susceptibles de edición y modificación, lo que facilita la recuperación de información con un mínimo esfuerzo de tiempo y coste.

7.- Gestión de mecanismos de control terminológico, tanto para la entrada de datos como para su recuperación. Puede tratarse de un tesauro, de un glosario o de un diccionario terminológico.

8.- Ayuda al usuario en todo momento, a través de mensajes y líneas de estado, especialmente durante el proceso de interrogación (interrogación asistida). En sistemas de recuperación en línea (teledocumentación), el sistema informa al usuario del tiempo de conexión, tareas ejecutadas, coste de la sesión, etc. Los mecanismos de ayuda al usuario, especialmente aquellos referidos a la evaluación y refinamiento de las búsquedas, son una de las principales áreas de investigación.

9.- Dependiendo de la configuración del sistema, éste puede ofrecer opciones de acceso multiusuario, niveles de seguridad, reorganización y recuperación de ficheros, etc.

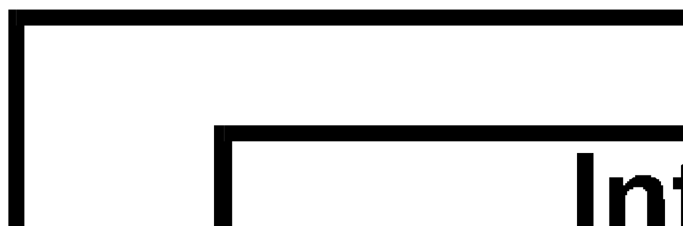


Fig.4.3. Esquema de un IRS (Information Retrieval System)

En general, las funciones anteriores se engloban en alguno de los seis subsistemas genéricos que debe ofrecer un STRID (LANCASTER y WARNER, 1993: 15-16):

1.- Subsistema de selección de documentos.
2.- Subsistema de indización.
3.- Subsistema de vocabulario.
4.- Subsistema de búsqueda.
5.- Subsistema de interacción hombre-máquina.
6.- Subsistema de comparación.

4.6. Consultas al STRID: las ecuaciones de búsqueda.

La búsqueda se realiza sobre un conjunto de términos introducido en el sistema por el usuario. Sin embargo, esta es la fase final de un proceso que comienza por el análisis de los documentos que se van a introducir, los datos que los conforman, y cómo estos forman una estructura de datos, como se ha visto en el capítulo correspondiente. Existen múltiples metodologías de diseño de

estructuras de datos, cada una con sus virtudes y sus defectos. La estructura de datos se plasma en el sistema informático en una estructura de base de datos, a partir de la cual se introducen los documentos. La organización de la información con vistas a su recuperación dentro de un sistema documental ha producido abundante bibliografía en la que pueden analizarse múltiples aspectos (TAGUE SUTCLIFFE, 1996).

4.6.1. El proceso de búsqueda.

El proceso de búsqueda de la información es más complejo de lo que pueda parecer en un primer momento, y tiene, con posibles variaciones (véase el panorama esbozado con anterioridad sobre la recuperación de información), las siguientes fases:

1.- Definición del tema de búsqueda.
2.- Selección de términos descriptivos sobre el tema de búsqueda. Consulta de tesauros o diccionarios, etc.
3.- Selección de las bases de datos sobre las que realizar la búsqueda.
4.- Formulación y ejecución de ecuaciones de búsqueda.
5.- Evaluación de los resultados. Replanteamiento, si procede, de las ecuaciones para ajustar la búsqueda.

Se debe considerar, además, que la definición del tema y de los términos para la búsqueda se hace en lenguaje natural, lenguaje que debe traducirse a la terminología empleada en la base de datos, y que los términos empleados en las ecuaciones son la traslación al lenguaje de interrogación de la base de datos de los términos originales, combinados mediante operadores que aplican lógica matemática. Tampoco hay que olvidar que la localización de esos términos dentro de la base de datos está en muchos casos sujeta a la estructura de campos existente. Lo anterior hace ineludible la utilización de lenguajes documentales para la correcta explotación de las bases de datos documentales.

En principio, debe aceptarse que la búsqueda incluye una serie de términos significativos, los cuales describen el área de interés para la recuperación de información. Estos términos mantienen entre ellos algún tipo de relación semántica. Esta relación debe reflejarse a través de la utilización de operadores que la reflejen, y por esta causa una ecuación de búsqueda se compone de dos tipos de elementos. En primer lugar, los términos que representan el tema de interés, y en segundo lugar, un conjunto de operadores que expresan la relación que deben mantener los términos entre sí. Este segundo tipo de elementos son el objeto de los siguientes párrafos.

4.6.2. Tipos de operadores utilizados en las ecuaciones de búsqueda.

La combinación de los términos necesarios para la recuperación de los documentos adecuados se basa en la utilización de unos operadores, especialmente los denominados booleanos, en honor de George Boole, que en 1847 publicó un trabajo titulado *The mathematical Analysis of Logic*, en el que sentó las bases de la denominada álgebra lógica o simbólica. La combinación de

los términos mediante los operadores crea una forma matemática, una ecuación, susceptible de ser tratada por medios informáticos.

Los operadores booleanos establecen relaciones entre los términos, y son O (operador de unión), Y (operador de intersección) y NO (operador de exclusión). Trabajan sobre el conjunto de los documentos, estableciendo subconjuntos con aquellos documentos que se ajusten a las condiciones fijadas en la ecuación.

La complejidad que rodea a los mecanismos de recuperación de la información ha hecho necesario aplicar otros operadores, que complementen a los booleanos. En primer lugar, los operadores de localización, referidos a la posición de los términos dentro de los documentos. Se basan en una hipótesis según la cual la cercanía entre dos términos puede significar una estrecha relación entre ellos. Pueden ser absolutos, cuando se establece que un término debe aparecer en un campo determinado, independientemente del resto de los términos incluidos en la ecuación, o relativos, cuando se establece que un término debe tener una posición referida a otro término incluido en la ecuación, por ejemplo en la misma línea, la misma frase, separados por "x" palabras, etc. Ambos pueden combinarse. También se encuentran los operadores de rango o intervalo, que establecen un intervalo, dentro del cual deben encontrarse los términos a recuperar, especialmente utilizados en la recuperación de datos numéricos y de fechas. Las dificultades que se desprenden de las características intrínsecas de los términos también afectan a la recuperación de la información. Se pueden plantear problemas con el uso de plurales de los términos, géneros o prefijos y raíces. Para intentar reducirlos, se emplean símbolos de truncamiento, cuya finalidad es indicarle al sistema de recuperación que se está buscando una cadena de caracteres dentro de un término, no un término en sí mismo.

Todos los operadores señalados pueden combinarse entre sí para crear ecuaciones complejas, considerando que siempre se establecen unas limitaciones por parte de los operadores. Normalmente, se ejecutan primero aquellas operaciones con los operadores más restrictivos, y su resultado se combina con los términos relacionados con operadores menos restrictivos. Los sistemas de gestión de bases de datos documentales incluyen en su documentación tanto los operadores disponibles y su uso, como las limitaciones y restricciones establecidas.

4.6.3. La recuperación mediante exploración.

La utilización de ecuaciones de búsqueda como principal herramienta en la recuperación de información plantea problemas, que los enfoques actuales de investigación pretenden resolver mediante la creación de mejores interfaces de usuario (INGWERSEN, 1992), que reflejen los mapas cognitivos de los usuarios, y permitan generar imágenes de los mismos que puedan ser utilizadas para interrogar los recursos de información. Interrogación que debería verse complementada por la capacidad de los sistemas de añadir funciones de exploración de los contenidos informativos, durante el proceso de recuperación. El fundamento de los procesos de exploración es la revisión y análisis de documentos por sus contenido, de forma que el sistema sea capaz de llevar al

usuario a otras informaciones similares o relacionadas, sin necesidad de formular nuevas ecuaciones de búsqueda.

4.6.4. Los resultados: análisis y modificación.

Una vez considerado todo lo expuesto, y realizada una búsqueda, es de rigor analizar su pertinencia. Una alta pertinencia se caracteriza por la recuperación de los documentos adecuados a la información que se deseaba obtener, evitando la introducción de otros que puedan distorsionar el conjunto, y cumplir los requisitos de una correcta recuperación de la información En sistemas de tratamiento y recuperación de información documental, ésta debe ser exacta, exhaustiva, precisa, oportuna, íntegra y significativa. Por contra, el silencio (falta de documentos) y el ruido (exceso de documentos, muchos de ellos no significativos) caracterizan a un bajo nivel de pertinencia.

Como resultado del análisis anterior, cabe plantearse la adecuación de la respuesta obtenida a lo esperado. Pueden darse dos situaciones, ante las cuales deben adoptarse diferentes acciones. En un primer caso, puede suceder que el resultado de la ecuación sea demasiado escaso. Entonces procede ampliarla, lo cual puede hacerse con la utilización de términos más genéricos, sinónimos, o ampliación de truncamientos. En un segundo caso, si el resultado obtenido es excesivamente amplio, deben utilizarse medidas contrarias: utilización de términos más específicos, reducción de truncamientos, etc.

Página en blanco

5. HIPERTEXTO E HIPERMEDIA.

Lecturas recomendadas: CANALS, 1990; ELLIS, 1990; GARCÍA MARCO, 1996; NIELSEN, 1990; SEYER, 1991; MARTIN, 1990; MOSCOSO y CARIDAD, 1991; CONKLIN, 1987; MCKNIGHT, DILLON y RICHARDSON, 1991; DÍAZ, CATENAZZI y AEDO, 1996; LANDOW, 1995; COVICE, 1994; RODRÍGUEZ DE LAS HERAS, 1991; LYNCH, 1994.

5.1. Orígenes del hipertexto.

Los términos hipertexto e hipermedia son ahora un lugar común al hablar de las nuevas tecnologías de presentación y acceso a la información, y se están revelando como una herramienta imprescindible en lo respecta tanto a la representación de documentos e información, como herramienta para procesos comunicativos y formativos en todos los ámbitos (CANALS, 1990). Sin embargo, y a pesar de su pretendida novedad, este enfoque de organización y acceso a la información tiene antecedentes desde la década de 1940 (GARCÍA MARCO, 1996). En tan temprana fecha, Vannevar Bush ideó un sistema de control, gestión y acceso a la documentación al que llamó MEMEX, y cuya idea principal era utilizar un principio de asociación de conceptos entre recursos informativos, de tal forma que el usuario pudiese acceder a estos recursos, independientemente de su tipo, simplemente utilizando la asociación de ideas. El sistema no llegó nunca se ser construido, por la dificultad técnica que suponía. Hubo que esperar a la década de 1960 para que Douglas Engelbart, un investigador del Stanford Research Institute, dirigió un proyecto de investigación para desarrollar máquinas, basadas en sistemas informáticos, que permitiesen aumentar la capacidad intelectiva humana, y gracias al cual se introdujeron conceptos técnicos sin los cuales no sería posible el hipertexto: sesiones interactivas, dispositivos señaladores, ventanas y escritorios... que permitían gestionar gran cantidad de información según criterios jerárquicos y asociativos. El final del proyecto supuso que gran parte del equipo de trabajo pasó a Xerox, donde desarrollaron los fundamentos de los sistemas operativos de interfaz gráfico tan extendidos en la actualidad (MacOS, Windows, X-Window...).

En la misma década que Engelbart, Theodor Nelson acuño el término hipertexto, al que definió como “escritura no secuencial”. Retomando el concepto de asociación de Bush, por otra parte completamente aceptado por los estudios sobre el comportamiento del pensamiento humano, Nelson trabajaba en un proyecto para introducir, organizar y recuperar toda la información del mundo en un sistema informático, utilizando el principio de la asociación de ideas y documentos, al que llamó Xanadu. En este sistema los documentos se relacionaban utilizando enlaces a fragmentos y citas comunes, en un entorno de red informática en el que podían participar numerosos ordenadores. El proyecto Xanadu, comenzado en la Universidad de Harvard, en la que no cuajó, pasó a ser desarrollado en Australia, y recientemente Nelson ha pasado a trabajar en varios proyectos de investigación en Japón, siempre en el ámbito del diseño y desarrollo de sistemas hipertextuales.

Como consecuencia directa de los proyectos enumerados, en la segunda mitad de la década de 1980 comenzaron a aparecer en el mercado gran cantidad de de aplicaciones y herramientas para ordenadores personales, que facilitaban la creación de sistemas hipertextuales a nivel personal, y que en los últimos años incorporan capacidades multimedia, generando de esta forma sistemas hipermedia. Las nuevas generaciones de STRID, así como los últimos desarrollos de SGBD incorporan mecanismos especializados para organizar la información utilizando, si el usuario lo desea, de forma hipertextual, incorporando mecanismos de acceso a diferentes tipos de información (gráfica, sonora...).

5.2. Conceptos de hipertexto e hipermedia.

5.2.1. Hipertexto.

Las ideas expuestas en el bloque anterior deben servir como una primera aproximación al concepto de hipertexto. Como señaló Nelson, el hipertexto se basa en la escritura, y en la lectura, no secuencial de los documentos (NELSON, 1990: 1-3). Esto supone la existencia, y la adopción, del criterio de asociación de ideas y conceptos como principio organizador del conjunto de documentos, del conjunto informativo (LANDOW, 1995). Este tipo de organización no puede reflejarse adecuadamente en material impreso, ya que en éste predomina la linealidad. Sin embargo, existe la posibilidad de superar esa linealidad, utilizando como herramientas sistemas informáticos capaces de establecer relaciones entre ficheros. Dos elementos clave en estos sistemas son los de enlace y nodo. Un nodo es un documento o elemento informativo, en soporte informático. Entre nodos pueden establecerse relaciones y asociaciones a través de enlaces. De esta forma, dos documentos que contengan conceptos relacionados pueden ofrecer acceso directo uno a otro, o a la parte de los mismos que sea pertinente. Por otra parte, es establecen varios caminos posibles para acceder a la misma información.

5.2.2. Hipermedia.

El panorama esbozado es perfectamente comprensible desde una perspectiva textual, en la que los nodos corresponden a documentos con información escrita. Sin embargo, la información y la documentación actual incorporan gran cantidad de componentes gráficos (estáticos o dinámicos), sonidos u otros complementos (MOSCOSO y CARIDAD, 1991). La evolución de la tecnología informática ha hecho posible que estos documentos no textuales sean incorporados en formato digital a los sistemas informáticos. Como consecuencia, los sistemas hipertextuales han evolucionado hasta disponer de las capacidades necesarias para incluir en los sistemas hipertextuales documentos e información con presentación gráfica y/o sonora. Esta inclusión, o expansión de las funciones de gestión de información a cualquier forma que pueda presentar la misma, manteniendo como criterio la asociación o relación de conceptos, es lo que caracteriza a un sistema hipermedia.

La creación de estructuras informacionales que integran hipertexto e hipermedia dan lugar a un nuevo concepto de documento electrónico dinámico, cuyas características ha resaltado Martin (MARTIN, 1990: xii):

> "The term hyperdocument refers to an electronic document which combines hypertext with diagrams, possibly sound, animation, or video, and possibly other types of software... with hyperdocument software we can create "intelligent documents" which guide their users, have built-in artificial intelligence or expertise, and can adapt themselves to user's needs. Many separate documents can be electronically interlinked."

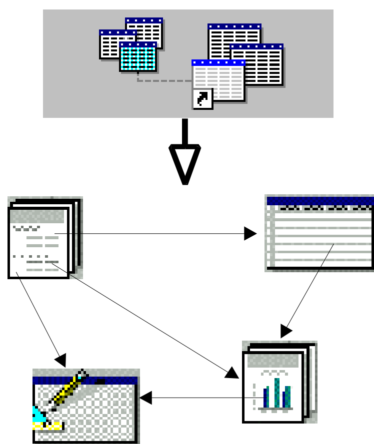


Fig.5.1. El hipertext/hipermedia se basa en el enlace entre informaciones de diferente tipo.

5. 3. Componentes de un sistema de hipertexto/hipermedia.

Para que sea posible una existencia real de los conceptos de hipertexto e hipermedia, deben utilizarse aplicaciones que sean capaces de crear los vínculos y asociaciones entre los documentos. Las aplicaciones ofrecen unos elementos particulares que facilitan la creación y navegación por las estructuras hipertextuales (CONKLIN, 1987; NIELSEN, 1990: 101-114):

1. Un conjunto de ficheros que contienen los documentos relacionados.
2. Ventanas de presentación de los documentos, las cuales son modificables en tamaño y posición.
3. Dispositivos señaladores, que facilitan la selección y el acceso a los documentos mostrados en las ventanas.
4. Punteros o enlaces, que generalmente utilizan una representación gráfica distinta a la del resto del material informativo, en forma de color, iconos, botones...
5. Herramientas de creación de enlaces y anotación de la navegación, lo que da al usuario la posibilidad de crear sus propias asociaciones y documentos.

Estas funcionalidades se integran en una herramienta que en el entorno hipertextual es conocida como “browser”, navegador o visualizador. El visualizador actúa como una interfaz, que muestra al usuario el contenido informativo de los documentos que selecciona, mediante la selección de enlaces. Suele completarse con la posibilidad de ejecutar búsquedas en el texto completo que contienen los documentos., y/o búsquedas más rígidas utilizando lenguajes clásicos de interrogación. La interrogación, sea de texto, imágenes o sonidos, suele realizarse a través de la ejecución de patrones, que representan una necesidad dada de información por parte del usuario. Además, una completa aplicación para este ámbito debería ser capaz de generar mapas gráficos de la estructura hipertextual, y utilizar estas representaciones para acceder directamente a los documentos deseados.

La visión que obtiene el usuario mediante el visualizador es una visión transparente, integrada, en la que no resulta complicado navegar de un documento a otro. Esta aparente facilidad no debe ocultar que los documentos pueden encontrarse en diferentes ficheros informáticos, e incluso en diferentes ordenadores, formando lo que se llama repositorio de información, que será tratado con más detalle en el próximo capítulo, por su relación con las bases de datos multimedia.

Los sistemas y estructuras de hipermedia pueden además incorporar inteligencia embebida, es decir, ser capaces de ejecutar otras aplicaciones o de tomar decisiones, acordes con la actividad desarrollada por el usuario, tanto en la utilización de los enlaces como en el acceso a los contenedores.

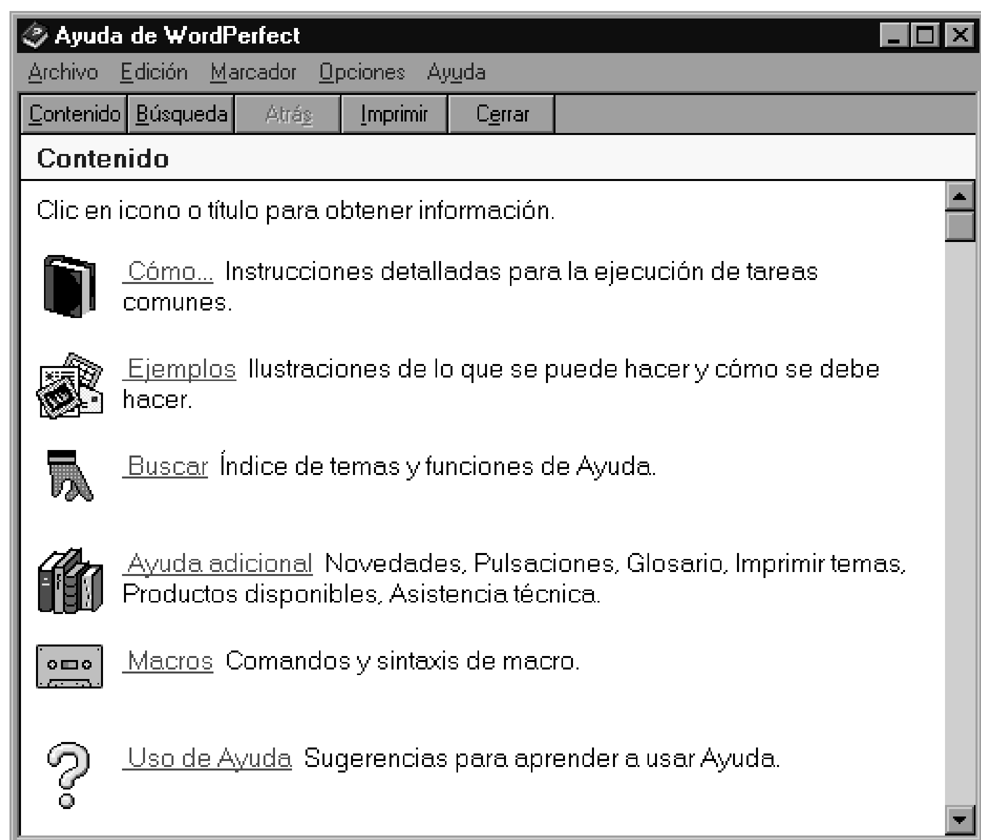


Fig.5.2. Ayuda de Windows 95: un sistema de hipertexto.

5.4. Tipos de sistemas de hiperdocumentos.

Las aplicaciones de los sistemas hipermedia pueden ser múltiples, y sólo se encuentran limitadas por la capacidad de la aplicación o herramienta elegida, y por los criterios fijados por los creadores y usuarios. Se han establecido algunas grandes categorías que engloban las posibles clases de sistemas (MARTIN, 1990):

1. Macroliterarios: trabajos y artículos relacionados con bibliografías, críticas, comentarios...
2. Colecciones de documentos: información organizada sobre un tema, de gran tamaño (servicios de información industrial y técnica, servicios bibliotecarios...)
3. Herramientas de exploración de problemas: organizar material e ideas desconectadas para estructurar una idea general.
4. Conjunto interrelacionados de información: informaciones diversas que son útiles si se asocian (modelos complejos, información policial...).

5. Documentos largos y complejos: manuales, regulaciones... (productos, metodologías, referencias legales, guías...).
6. Documentos inteligentes: combinado con inteligencia artificial mediante sistemas expertos (Diagnósticos, autoenseñanza, ayuda sensible al contexto...)

La anterior divisiión no es obstáculo para poner en relación, en un sistema informático, una aplicación de hipertexto con otras aplicaciones, de cuya interacción puedan beneficiarse ambas, como pueda ser un sistema de gestión de bases de datos, aplicaciones ofimáticas, sistemas de apoyo a la decisión, ayudas en línea o sistemas expertos.

5.5. Navegación y exploración de un sistema hipertextual.

La utilización de un sistema hipertextual por parte de un usuario se basa en la navegación o exploración de los contenidos informativos del mismo. (MCKNIGHT, DILLON y RICHARDSON, 1991: 75-86). Exploración o navegación que ofrece diferencias respecto a la lectura tradicional de un documento, ya que el documento hipertexto posee características distintivas, como la duración temporal, la situación espacial, la variabilidad de la presentación o la reutilización del documento o de sus partes en la generación de nuevos documentos (LYNCH, 1994).

La exploración de un espacio de información como el que configura un sistema hipertexto/hipermedia puede provocar problemas en el usuario, como el citado desbordamiento cognitivo. El propio mecanismo de exploración/navegación se basa en el paso entre conceptos asociados, transición establecida y controlada por el usuario final (COVICE, 1994). Por lo tanto, los sistemas deben ofrecer mecanismo de apoyo a la navegación, en dos áreas:

1.- Revisión histórica del proceso de exploración.
2.- Situación del usuario dentro del espacio informativo.

Los sistemas deben ofrecer al usuario mecanismos para controlar en todo momento el proceso seguido, desde el comienzo, hasta llegar al documento actual. De esta forma se puede limitar el impacto de la pérdida del objetivo inicial. Una segunda acción es la posibilidad de utilizar mapas de la estructura del sistema de hipertexto, que permitan al usuario conocer su situación en todo momento. De esta forma se complementa el acceso por exploración con el posible acceso directo al documento o nodo deseado. Estos mecanismos de apoyo pueden implementarse directamente, o bien utilizar metáforas (bi o tridimensionales), representativas de las acciones y decisiones del usuario (NELSON, 1990: 127-137).

5.6. La recuperación de información en un sistema de hipertexto o hipermedios.

El acceso al documento, en el contexto de los sistemas de recuperación de información y de bases de datos documentales, se realiza, generalmente, mediante la composición y ejecución de expresiones (ecuaciones) de búsqueda. Sin embargo, y como se ha señalado con anterioridad, en principio organizador de las estructuras hipertextuales es la asociación. Esto implica que el mecanismo de acceso y selección de la información relevante es diferente al "tradicional", y que los processos mentales desarrollados por el usuario serán diferentes en varios aspectos.

En el acceso y la recuperación de información en un sistema de hipertexto o hipermedios se han diferenciado seis grandes categorías de actividad (ELLIS, 1990):

1. Inicio: fase inicial de búsqueda de información.
2. Encadenado: proceso de navegación siguiendo las secuencias entre documentos o conceptos pertinentes, utilizando los enlaces existentes en los documentos.
3. Búsqueda por visualización: también llamada búsqueda semi-directa. Consiste en consultar listados de referencias o enlaces que el sistema debe ofrecer al usuario, previo al acceso al documento.
4. Diferenciación: empleo de diferencias entre varias fuentes, y sus tipos, como criterio de filtro y calidad de los documentos.
5. Monitorización: control del desarrollo de un área de conocimiento utilizando como referencia fuentes seleccionadas.
6. Extracción: uso sistemático de una fuente como herramienta de identificación de información pertinente.

Las categorizaciones establecidas demandan del sistema que soporta la estructura hipertextual avanzadas funcionalidades: recuperación en texto completo, compilación de listados de referencia, mecanismos de creación de mapas de navegación... De ello se puede deducir que la capacidad de recuperación de información en un sistema de estas características está íntimamente ligado con las herramientas de soporte que ofrece, no sólo de organización y de visualización.

5.7. La creación y organización de hiperdocumentos.

Los sistemas de hipertexto e hipermedia ofrecen al usuario la posibilidad de navegar, utilizando criterios de asociación, entre gran cantidad de documentos. En numerosas ocasiones, y dependiendo del tipo de sistema, esas asociaciones han sido creadas por terceros, según un conjunto de criterios lógicos y comprensibles, o al menos así lo han intentado. Sin embargo, la bibliografía ha recogido la existencia de problemas cuando un lector se enfrenta a una estructura de información que utiliza el hipertexto como principio organizativo (RODRÍGUEZ DE LAS HERAS, 1991).

Entre esos problemas deben citarse la posible desorientación del usuario, causada por el desconocimiento de la estructura de la información y de los medios existentes para navegar por ésta; el desbordamiento cognitivo, provocado por la dificultad en mantener el mismo nivel de concentración en el proceso de navegación que en la comprensión de la información a la que se accede; por último, los posibles errores en la interpretación de la relación implícita en los enlaces y nexos que unen los documentos.

Para crear hiperdocumentos es preciso respetar ciertas normas que eviten el amenazador "hipercaos". El hiperdocumento se divide en unidades de información, cada una de las cuales se encuentra en un envoltorio o contenedor, organizándose mediante un conjunto de criterios que pueden combinar la jerarquía con la asociación, de forma tal que la organización de las unidades sea claramente visible para el usuario. El factor de éxito en la construcción de hiperdocumentos es la división en nodos y los criterios pertinentes adoptados. Sin embargo, no debe olvidarse en ningún momento que el objetivo final es la comprensión activa y provechosa por parte del usuario, en un proceso comunicativo diferente a la lectura secuencial tradicional. Esto exige claridad de estructura, de organización de ideas, de lenguaje, de gráficos e ilustraciones y de mecanismos de navegación, de tal manera que el hiperdocumento resulte ser mucho más visual que el libro tradicional. Todas estas cuestiones hacen inevitable un riguroso estudio del usuario: quién es, porqué lee, qué sabe, que comprende, qué le resulta de valor...

Se han definido varias aproximaciones para la construcción de hiperdocumentos (SEYER, 1991: 93-117; DÍAZ, CATENAZZI y AEDO, 1996: 146-163). De esta forma, el diseñador puede optar por utilizar aproximaciones basadas en tabla de contenidos, jerárquico o de índices (en todos ellos se seleccionan opciones tomadas de menús que se ofrecen al usuario), mediante hipermapas (gráficos que representan la estructura del conocimiento almacenado en el hiperdocumento), visitas guiadas (especialmente para usuarios no expertos), tablas de decisión (según las opciones seleccionadas por el usuario se accede o ejecutan diferentes procesos), o plenamente empírica (basada en al experiencia del usuario y del diseñador, y en la observación del comportamiento del usuario). Por supuesto, estas aproximaciones no son excluyentes, si no que pueden combinarse.

Para terminar, cabe realizar una última, pero no por ello menos importante, consideración. Por sus propias características, un verdadero sistema hipertextual no debería ser un sistema cerrado, sino un sistema abierto y dinámico. Esto supone la necesidad de establecer unas normas de actualización y mantenimiento de la información introducida, de tal forma que se asegure la consistencia del conjunto, independientemente de su variabilidad espacial y temporal.

6. LAS BASES DE DATOS MULTIMEDIA.

Lecturas recomendadas: CHORAFAS, 1994; MCKNIGHT, DILLON y RICHARDSON, 1991; LÓPEZ YEPES, 1993; NICHOLLS y RIDLEY, 1996; KOEGEL, 1994; FIDEL *et alii*, 1994; BIANCHI *et alii*, 1996; APERS, BLANKEN y HOUTSMA, 1997; AZORÍN y LÓPEZ, 1994; DÍAZ, CATENAZZI y AEDO, 1996; FONG y SIU, 1996; MEGHANI, RABITTI y THANOS, 1991; OZKARAHAN, 1995.

6.1. Multimedia y nuevos documentos.

El término multimedia ofrece dos problemas. En primer lugar, el entorno, mercado o ámbito que delimitan a multimedia todavía no ha sido bien definido, y sus límites resultan difusos. en segundo lugar, multimedia es un término que puede significar diferentes cosas, según quién lo utilice. Originalmente, el término multimedia, a finales de la década de 1970, resultaba ser la integración de voz, texto, datos y gráficos, a los que habría que añadir, actualmente, los gráficos interactivos, la imagen en movimiento, la composición de documentos electrónicos y la realidad virtual (KOEGEL, 1994).

La aparición de aplicaciones multimedia sólo ha sido posible gracias al desarrollo de entornos gráficos de usuario para el aprovechamiento de los computadores. La aparición de estos entornos gráficos (Macintosh, Windows, X-Window, NextStep...), basados en la presentación de información en paneles o ventanas, posicionables en la pantalla, y en la utilización de representaciones pictóricas (iconos) como indicadores de elementos, así como de menús de opciones rápidas seleccionables por el usuario, ha modificado en profundidad la manera de presentar y de interactuar sobre los elementos informativos.

Uno de los problemas a resolver ha correspondido al almacenamiento y acceso a la información de tipo gráfico. A lo largo de los últimos años han proliferado gran cantidad de formatos de almacenamiento para gráficos (PICT, TIFF, GIF, JPEG y otros), que han hecho necesario la creación de filtros y programas de transferencia entre unos y otros. Afortunadamente, se comienzan a

establecer ciertos niveles de compatibilidad, con vista a su utilización en bases de datos, a través de metaficheros gráficos, que faciliten el acceso a todos los formatos, independientemente de cual se trate y de las aplicaciones que los manipulen.

Por último, el auge de multimedia debe gran parte de su éxito a la popularización de los medios ópticos y magneto-ópticos de almacenamiento masivo. La gran cantidad de espacio necesaria para el almacenamiento de ficheros de imagen de calidad adecuada, y el acceso y manipulación de éstos, dentro de un tiempo de respuesta prudencial, ha recibido una respuesta adecuada con el desarrollo de las últimas generaciones de CD-ROM y de lectores de alta velocidad. Tampoco puede pasarse por alto que la gran cantidad de espacio de almacenamiento requerida por imágenes y grandes volúmenes de información textual demanda, a pesar de los medios ópticos de almacenamiento, la utilización de técnicas de compresión y descompresión de ficheros de alta velocidad, que en su mayor parte se realizan mediante software específico, de forma que se combine un aceptable nivel de ocupación de recursos con una garantía de calidad elevada, en lo que respecta al contenido.

Los párrafos anteriores cubren cuestiones eminentemente relacionadas con los soportes la información. Sin embargo, un sistema multimedai debe contemplarse desde dos niveles. En primer lugar un nivel intencional, superior, en el que se recogen los constructos del conocimiento humano que son creados o representados. En segundo lugar, un nivel extensional, inferior, que engloba la organización de los datos. Los conceptos de procesamiento de la información multimedia se distribuyen en ambos niveles. Los nuevos desarrollos en las bases de datos multimedia extienden las concepciones de los sistemas de información y de sus usuarios más allá de la simple integración. Técnicas de inteligencia artificial y sistemas expertos, mecanismos de representación y recuperación de información gráfica, desarrollo de nuevas interfaces de usuario, integración general de hipertexto e hipermedia, aplicaciones en el campo de la edución y la formación, etc., (APERS, BLANKEN y HOUTSMA, 1997): Los STRID que se impondrán en el mercado de pocos años incorporarán al tratamiento documental tradicional todas estas herramientas.

6.2. La integración de multimedia en las bases de datos.

La aparición y desarrollo de la tecnología multimedia ha revolucionado el concepto tradicional de base de datos, entendidas como elementos de información textual y numérica, los cuales a su vez eran organizados de acuerdo a un conjunto preestablecido de normas. Estos componentes se ven enriquecidos por la posibilidad de completar los objetos a los que representan mediante elementos gráficos y/o sonoros, en un entorno integrado. El usuario obtiene entonces una representación compleja, múltiple, de los objetos del mundo real, y de sus categorías de información, que son representados en la base de datos. El ambiente de un sistema de gestión de bases de datos multimedia integra texto, datos, video, imagen (estática y dinámica) y sonido.

La integración de diferentes tipos de información en un único documento pone de relieve la importancia de las tareas de representación del contenido informativo, especialmente la indización de imágenes y documentos sonoros. Los mecanismos de recuperación de información de estos sistemas siguen utilizando, (excepto aquellos más punteros, todavía en fase de investigación) términos como elemento de formulación de requerimientos. La utilización de complejos mecanismos y normas de indización de estos documentos merece cada vez en mayor volumen la atención de los investigadores (WALKER y JONES, 1994; AZORÍN Y LÓPEZ, 1994), por lo menos hasta que se difundan mecanismos de recuperación basados en patrones gráficos o técnicas similares.

En una base de datos multimedia se está trabajando con un marco que no sólo incorpora los datos. Esos datos pueden tener variabilidad espacial y temporal. Por lo tanto, un documento introducido en una base de datos multimedia es una composición temporal, en la cual hay que introducir los diferentes tipos de datos, tanto como las relaciones de configuración y temporales existentes entre ellos. Los datos deben estar sincronizados, controlando tanto su estado como su comportamiento.

6.2.1. Los documentos multimedia.

Los documentos multimedia se forman mediante la agregación de objetos de diferentes tipos, objetos que pueden ser de gran tamaño, y llegan a mostrar gran variedad en lo referido a estructura y representación. La representación del documento multimedia busca comunicar de forma efectiva no sólo el documento y su contenido, sino también la semántica, los conceptos subyacentes y la relación entre ellos. Una comunicación de información que siga estas normas tiene asegurado un adecuado nivel de calidad. Por lo tanto, resulta ineludible fijar como uno de los enfoques básicos el análisis de las características de los usuarios humanos del sistema, para lo cual deben analizarse los patrones de comportamiento del usuario en el acceso y comprensión de las estructuras informativas (BIANCHI *et alii*, 1996):

Sirva el párrafo anterior como introducción para establecer una primera conclusión: en el momento en el que se comienza a diseñar e implantar una base de datos multimedia, el objeto de trabajo ya no resulta ser el registro tradicional: el usuario está tratando con un documento electrónico. Este ha sido creado procesando los diferentes tipos de información (texto, gráficos, imágenes, datos, imágenes animadas, sonidos) presentes en un almacén o repositorio de información. A esto hay que añadir que tras largos años de implantación y desarrollo de las tecnologías de bases de datos, las organizaciones y empresas disponen de grandes volúmenes de información almacenadas en bases de datos, generalmente en diferentes plataformas y aplicaciones, distribuidas en diferentes localizaciones geográficas. Estas circunstancias exigen que la implantación y explotación de bases de datos multimedia exija dos elementos:

1. La disponibilidad de bases de datos avanzadas, en un entorno de fácil utilización por parte del usuario.
2. El establecimiento de redes de telecomunicaciones de alta velocidad.

6.2.2. Los sistemas de gestión de bases de datos multimedia.

De hecho, es la heterogeneidad de los tipos de información que son necesarios en la actualidad unas las razones que ha favorecido, por parte de la industria y los usuarios, el desarrollo de sistemas de gestión de bases de datos multimedia, que han sido llamados también “gestores de información hipermedial” (DÍAZ, CATENAZZI y AEDO, 1996: 174). Las aplicaciones SGBD tradicionales ofrecían limitaciones en aspectos como el acceso complejo a datos, la transferencia de datos con otros sistemas, o la inexistencia de adecuados interfaces de usuario. Como respuesta, se tiende a diseñar e implementar nuevos SGBD que sean capaces de utilizar "inteligentemente" los datos disponibles, e integrar las viejas y las antiguas aplicaciones de forma no traumática. Una base de información hipermedial tiene varios componentes:

1. Base de presentación: parámetros a aplicar para mostrar la información al usuario.
2. Base de estructura: visión lógica del hiperdocumento, según un modelo.
3. Base de contenido: conjunto de documentos que se integran en el hiperdocumento.
4. Base de utilización: información sobre hábitos y comportamiento de cada usuario.

En este mismo sentido, la concepción de una base de datos multimedia, en su modelo conceptual (correspondiente al esquema conceptual definido por ANSI/X3/SPARC), debe cumplir dos fases (CHORAFAS, 1994: 312):

1. Cognición, centrado en cómo reconocer el mundo real, sus entidades y relaciones.
2. Modelización, centrado en cómo representar los conocimientos obtenidos en la fase anterior, de manera que sean manipulables por la máquina.

A pesar de ser la integración de bases de datos heterogéneas una de las razones del desarrollo de sistemas de gestión de bases de datos multimedia, la industria todavía no ha establecido todavía estándares de formato entre ellas, por lo que se repite la misma situación que en momentos anteriores, entre ficheros de base de datos correspondientes a aplicaciones como dBASE, Paradox, etc. Aunque las aplicaciones tradicionales ya ofrecen "puentes" para compartir sus bases de datos, por el momento esto no es posible en lo que respecta a las bases de datos multimedia. Sin embargo, están comenzando a aparecer en el mercado nuevas aplicaciones que, sin cumplir los requerimientos clásicos de un SGBD, pueden generar documentos multimedia, tomando como base información y datos contenidos en otros tipos de fichero. Los límites entre la base de datos "real", y el documento multimedia resultante "virtual", son difíciles de establecer en el estado actual de la tecnología. Si puede decirse que es posible establecerse un algoritmo que define el proceso de formateo y composición del documento, en dos niveles:

1. Jerarquía lógica del contenido del documento.
2. Jerarquía física del documento formateado.

Esta es la premisa que define la norma ODA (Open Document Architecture) de ISO. En lo que respecta a otros estándares, la norma ISO que define el EDI (Electronic Document Interchange, y su versión EDIFACT), y el protocolo ANSI X.12 están más orientados a la estructura del documento que a sus posibles aspectos multimedia, lo que por el momento dificulta su aplicación real.

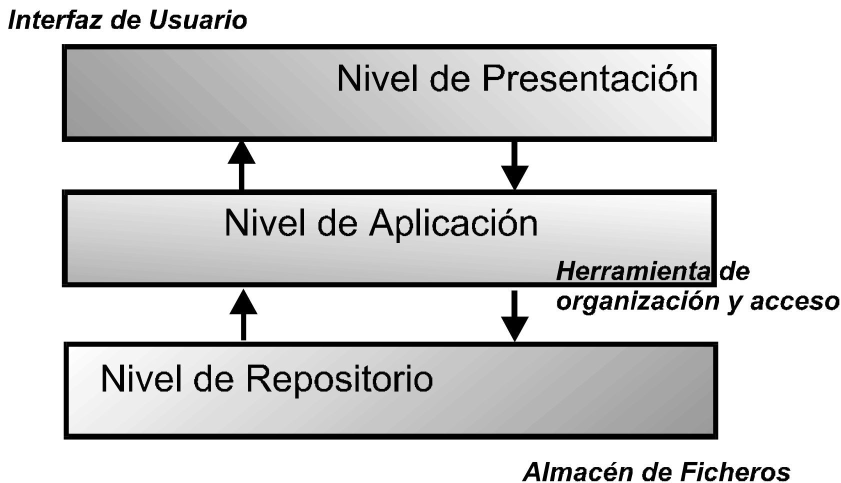


Fig.6.1. Los tres niveles de un sistema de bases de datos multimedia.

6.3. La metodología en una base de datos multimedia.

La creación e implantación de un sistema de bases de datos multimedia no puede separarse de la adopción de una metodología adecuada para ello. De la misma forma que en los sistemas "tradicionales" de bases de datos, el enfoque a utilizar es el basado en un sistema integrado de información:

1. Definición de las tareas y objetivos a cumplir por la unidad informativa.
2. Desarrollo de un metamodelo y de una metodología para el desarrollo de un sistema de información estratégica.
3. Inventario de los elementos de tecnologías de la información disponibles.
4. Inventario de las herramientas y de los repositorios de información existentes.
5. Creación de un diccionario de metadatos y de su metodología.
6. Consolidación de las conclusiones obtenidas, a través de metamodelos.

La utilización de este tipo de bases de datos, en lo que se incluyen representaciones complejas de la realidad, hace necesario la utilización de dos niveles de organización y de descripción. En primer lugar, una nivel de

conocimiento, de metadatos, conceptual y difuso. En segundo lugar, un nivel de datos, concreto y analítico. El concepto clave de las nuevas bases de datos, a partir de la presente década, será "significado", superando el clásico "dato" o "información".

6.4. Documentación y bases de datos multimedia.

En lo que respecta al ámbito documental, las bases de datos multimedia facilitarán el enriquecimiento de la representación de la información (LÓPEZ YEPES, 1993). El documento no se reducirá a la utilización de unas categorías descriptivas, complementadas con la aplicación de un lenguaje documental. La inclusión de nuevos elementos descriptivos de los documentos, como una imagen, gráficos, o complementos sonoros, requerirán la experimentación, el desarrollo y la utilización de nuevos mecanismos de recuperación de información, de los cuales ya se están utilizando, a determinados niveles, algunos de ellos, como el reconocimiento e identificación de imágenes según esbozos de líneas maestras. La aplicación de representaciones semánticas de la información, seguramente basada en el enfoque orientado a objetos, que incluye tanto los datos como las acciones a ejecutar sobre los mismos o sobre terceros, será uno de los paradigmas a emplear en este entorno.

Una cuestión clave, en los nuevos sistemas de bases de datos multimedia, es el tratamiento de la imagen de los documentos, tanto en lo referido a su almacenamiento, como en lo referido a su procesamiento. De hecho, esta es una de las mayores áreas de expansión de la informática documental en los próximos años, sirviendo como indicador el auge que están alcanzado los llamados SGD (Sistemas de Gestión Documental). Los documentos, en cuyo origen suelen encontrarse en soporte papel, son introducidos en el sistema a través de algún mecanismo de captura, generalmente un escáner. Se almacenan como gráficos (en cuyo caso se pierde la posibilidad de efectuar búsqueda a texto completo), o como texto, mediante la utilización de una aplicación de tipo OCR (Optical Character Recognition), que permite integrarlo como un texto perfectamente manipulable y accesible, desde una perspectiva documental. Evidentemente, una aplicación que incorpore la segunda posibilidad, o ambas, posee una importante ventaja sobre los meros almacenes de imágenes de documentos.

Esta última cuestión pone de manifiesto que un adecuado sistema de gestión de bases de datos multimedia debería ofrece herramientas para buscar, recuperar, manipular, ordenar y organizar los textos y las imágenes en un entorno de formato libre, de tal forma que pueda utilizar documentos de muy diverso formato, tamaño y disposición (NICHOLLS y RIDLEY, 1996; OZKARAHAN, 1995). El sistema de gestión de bases de datos multimedia debería ofrecer lenguajes de descripción de documentos avanzados, permitir el diseño de estructuras de datos muy flexibles, y ofrecer mecanismos de búsqueda altamente efectivos. Como corolario, un sistema de este tipo debería (CHORAFAS, 1994: 338-339):

1. Encontrar rápidamente la información multimedia, buscando cualquier número de objetos referidos a un contenido de información, en un

contexto de texto completo (y, a poder ser, en un contexto totalizador de la información)
2. Conectar documentos según su contenido informativo, relacionando extractos de información relacionada de varios documentos, en uno nuevo.
3. Facilitar el acceso instantáneo a los ficheros pertinentes, incluyendo los mecanismos necesarios de compresión y descompresión para la manipulación de aquellos.
4. Conocer y manipular, por parte del usuario, la estructura misma de la información.
5. Crear relaciones entre grupos de elementos informativos.

Los documentos pueden encontrarse en una única localización, en un computador local, o bien ser el resultado de la integración de datos y representaciones dispuestos en ordenadores dispersos espacialmente. Esta última situación obliga a tratar el asunto de las bases de datos distribuidas. Se está hablando de una única base de datos a nivel lógico, pero de diferentes bases de datos a nivel físico. Esto supone que las aplicaciones deben acceder a diferentes tipos de información, en diferentes estructuras, a través de redes de ordenadores, en entornos sumamente heterogéneos. La necesidad que se deriva del panorama esbozado es la presencia de un diccionario de datos, repositorio general a través del cual se disponen las definiciones estándares de los objetos presentes en toda la extensión de la base de datos.

6.5. Hipermedia.

Un desarrollo de las bases de datos multimedia distribuidas, combinadas con el hipertexto, es el servicio hipermedia (*vid. supra*). Hipermedia es un enfoque flexible y muy eficiente, para la gestión de información, que permite la navegación a través de la información distribuida. Su principio básico es la posibilidad de navegar, de forma interactiva, en una red formada por nodos conectados entre sí. El hipertexto sería un subconjunto del hipermedia, principalmente referido a información textual. La navegación se realiza saltando por la información, cualquiera que sea su tipo, de forma no secuencial, a través de links o punteros, establecidos según ciertos principios inteligentes, e inteligibles. De esta forma, un sistema hipermedia mostraría todo tipo de información almacenada en diferentes bases de datos y en diferentes formatos, de manera completamente transparente para el usuario.

Una hiperbase debería cumplir los siguientes requisitos (DÍAZ, CATENAZZI y AEDO, 1996: 185-191):

1. Toda información es enlazable.
2. El hiperdocumento es dinámicamente modificable.
3. Es necesario guardar información sobre la presentación física de los contenidos.
4. Se debe dar soporte a todos los tipos de acceso a los sistemas hipermediales.

5. El hiperdocumento debe ser abierto y puede estar físicamente distribuido.
6. El sistema debe incluir mecanismos de recuperación ante fallos.
7. Debe permitirse el acceso concurrente.
8. Hay que preservar la seguridad de la información.
9. Hay que posibilitar el trabajo en cooperación.
10. El hiperdocumento puede estructurarse siguiendo uan determinada jerarquía.
11. Existen contenidos virtuales enlazados dinámicamente.

7. LA TELEDOCUMENTACIÓN.

Lecturas recomendadas: ABOBA, 1993; WALKER y JONES, 1993; LANCASTER y WARNER, 1993; AMSTRONG y LARGE, 1988; BASCH, 1993; BATES, 1987; CHAMORRO y GONZÁLEZ, 1992; DVORAK y ANIS, 1992; FUNDESCO, 1986; HAWKINS *et alii*, 1985; LIZASOAÍN, 1992; MEADOW, 1992; COX, 1991; CONVEY, 1992.

7.1. Concepto de teledocumentación.

La teledocumentación, que consiste en el acceso y la consulta de bases de datos existentes en una localización remota, lejana al usuario, mediante la utilización de un ordenador conectado a una red de telecomunicaciones, ha sido considerada durante largo tiempo como el punto de referencia inexcusable en la informática documental (WALKER y JONES, 1993). Sin embargo, resulta necesario situar en su justo lugar la teledocumentación. Esta no sería posible sin la previa tarea de diseño y creación de bases de datos documentales, y sin el desarrollo de diferentes técnicas de recuperación de información.

Es conveniente, en este punto, establecer dos hipótesis de trabajo sobre la teledocumentación. En primer lugar, una de las bases de la teledocumentación es la limitación física del espacio. Tras el nacimiento de numerosas organizaciones gubernamentales, a diferentes escalas, desde inicios del presente siglo, así como la proliferación de la información científica, la cantidad de documentos generados por los mismos, que pueden ser de interés para un usuario en un momento dado, es enorme. Difícilmente una biblioteca, archivo o centro de documentación pueden ofrecer acceso directo al documento, ya que les es imposible almacenarlos todos físicamente. Y en segundo lugar, aunque fuese posible, el coste económico que supondría la adquisición y mantenimiento de la documentación resultaría inasequible. Para satisfacer a sus usuarios debe ofrecer algún tipo de mecanismo que ofrezca acceso a ese gran volumen de documentación, a un coste razonable.

La solución al problema pasa por que un tercero ponga a disposición de otros muchos usuarios los documentos, y algún mecanismo para facilitar la localización y acceso a los que sean de interés para el usuario. Lo más probable es que ese repositorio de información se encuentre en alguna situación lejana, lo que implica la necesidad de redes de comunicaciones para facilitar el acceso al mismo, así como la utilización de algún tipo de mecanismo que, utilizando esas redes, consulte los contenidos del fondo. De esta manera se llega a la definición clásica de teledocumentación, considerada como la unión de informática, telecomunicaciones y documentación.

El concepto de teledocumentación se ha englobado, por influencia anglosajona, bajo los términos “online searching”, término que podría ser válido en la década de 1970. En el momento actual, con la expansión de las redes de telecomunicaciones y de las bases de datos en soporte CD-ROM, el proceso de búsqueda resulta ser en línea, “on line”, el 90% de las ocasiones, lo que hace preferible utilizar el término teledocumentación, con unas características ya definidas, frente al término “online”, basado en consideraciones técnicas ya superadas.

7.2. Los actores de la teledocumentación.

La teoría tradicional sobre teledocumentación (LANCASTER y WARNER, 1993) establece la existencia de cuatro actores en el proceso teledocumental:

1. Productores de bases de datos: son los diseñadores y creadores de las bases de datos, que posteriormente son accesibles al usuario. Estos creadores desarrollan un producto, según diferentes estudios de mercado y de necesidades de información, seleccionado el área de cobertura, los fondos incluidos, y dan soporte al desarrollo y crecimiento posterior de la base de datos.
2. Distribuidores o "host": este segundo actor es el encargado de situar la base de datos a disposición de los usuarios interesados, actuando como intermediario entre el productor y el usuario. Establece una serie de convenios con los productores para ofrecer sus bases de datos, a través de contratos con los usuarios. Normalmente suelen disponer de una compleja estructura informática para hacer frente a numerosas consultas simultáneas por parte de los usuarios.
En numerosas ocasiones, los host son también productores de bases de datos, lo que puede abaratar sus servicios. Por contra, los productores suelen licenciar sus bases de datos a varios distribuidores, con la finalidad de favorecer la competencia y obtener mejores resultados.
4. Redes de comunicaciones: las redes de comunicaciones son la infraestructura técnica que permite el establecimiento de comunicación entre el distribuidor o host y los usuarios finales. Dependiendo de la situación y de la normativa existente en cada país, la tipología puede ser muy variada, tanto en el ámbito económico como en el ámbito técnico.
5. Usuarios finales: son los usuarios y demandantes de la información almacenada en las bases de datos mantenidas por los host. Disponen de una terminal informática conectada a la red de comunicaciones pertinente,

lo que les permite acceder a las bases de datos ofrecidas por el distribuidor, tras el pertinente contrato de ambos servicios.

7.2.1. El papel de los soportes ópticos.

La teledocumentación ha sido puesta en entredicho en el cambio entre la década de 1980 y la década de 1990. La popularización de las bases de datos documentales en soportes ópticos, especialmente en CD-ROM, a un precio moderado, hizo pensar en ocasiones en una decadencia de los distribuidores tradicionales. Sin embargo, unos pocos años han bastado para poner las cosas en su justo sitio. No hay que ver ambas posiciones como enfrentadas, sino como complementarias. El CD-ROM ofrece un acceso local, a un coste módico, a bases de datos muy completas, que sin embargo ofrecen el problema de la actualización. No es posible una actualización inmediata de las bases de datos en CD-ROM, lo que sí es ofrecido por la teledocumentación. Como puede imaginarse, en áreas donde la actualización y la novedad en el estado de la información son decisivas, poco puede aportar el CD-ROM, como en el caso de una agencia de noticias. Sin embargo, en lo que se refiere a archivo, el CD-ROM posee una ventaja innegable frente a la teledocumentación (LIZASOAÍN, 1992). Ambos deben considerarse bajo la óptica de las necesidades del usuario y del tipo de información, más que sobre las características de la tecnología utilizada.

7.2.2. Otros servicios de información en línea.

El mundo de los servicios de información en línea resulta ser sumamente dinámico y cambiante. En épocas recientes se ha visto una rápida aparición y caída de algunos sistemas, como el teletexto (CHAMORRO y GONZÁLEZ, 1992), o la generalización de sistemas de información y comunicación interpersonales, como los BBS (Bulletin Board Systems) (ABOBA, 1993). El teletexto, una iniciativa europea principalmente apoyada por Francia, ha tenido una vida y resultados efímeros, excepto en el caso del Minitel galo. Un insuficiente nivel de definición de la imagen, la lentitud en el acceso a la información y el elevado coste han sido definitivos en su fracaso, a pesar de ser una opción válida, en principio, para popularizar el acceso a la información mediante redes de telecomunicaciones. Muy diferente ha sido el caso de los BBS, los cuales se han desarrollado como foros electrónicos de acceso e intercambio de información, principalmente en los Estados Unidos de América, con empresas señeras como Compuserve o American On Line (AOL). Compuserve ofrece acceso directo a gran cantidad de bases de datos sobre todo tipo de cuestiones, incluyendo revistas con artículos a texto completo de variados ámbitos, información económica y comercial... La facilidad de acceso, la gran cantidad de información de todo tipo contenida en ellos, la posibilidad de intercambiar información con otros usuarios, los sistemas de mensajería electrónica, y la incorporación de algunos servicios de acceso a bases de datos, a través de las propias BBS, y el bajo coste, similar al de una llamada telefónica, así como la utilización como terminal de cualquier ordenador equipado con modem, han sido factores de éxito que han favorecido, casi sin transición, la entrada en el mundo de Internet.

7.3. Herramientas del usuario final.

El documentalista, como principal usuario final, e intermediario de otros posibles usuarios, debe disponer de un equipo que le permita acceder al distribuidor con las adecuadas garantías. Deberá disponer de un ordenador, que cumple las funciones de terminal del host, en el momento en el que entre en comunicación con éste. Sin embargo, los actuales ordenadores no vienen preparados, en su mayor parte, para la conexión a una red física de comunicaciones, por lo que resulta necesario adecuar el equipo a tal fin.

Para utilizar los sistemas de información electrónica disponibles a través de redes de comunicaciones hay que completar el ordenador con dos tipos de componentes, tanto de hardware como de software (DVORAK y ANIS, 1992). En primer lugar, es necesario instalar y configurar un modem (modulador/demodulador). La función del modem es la de actual como interlocutor entre el ordenador y la red de comunicación telefónica. Por lo tanto, deberá convertir la información del ordenador (digital) en información capaz de viajar por la red de comunicaciones (analógica), y viceversa. En ocasiones el ordenador puede estar conectado a algún tipo de red de comunicaciones entre ordenadores que hará innecesario este tipo de adminículo, en cuyo caso deberá ser sustituido por una tarjeta de red, que cumplirá el mismo cometido, aunque sin la necesidad de traducir entre analógico y digital.

Para acceder al host será necesario instalar un software o aplicación capaz de interactuar con el modem y presentar la información en nuestra máquina local de forma inteligible, para que cumpla su función de terminal. Se le suele denominar software de comunicaciones y/o terminal. Este software es el encargado de establecer la comunicación con el host, representar en la pantalla del usuario la información que se le envía, y enviar a su vez al host, las órdenes emanadas del usuario. Las características de este software pueden variar enormemente, dependiendo del gusto de usuario, de las facilidades de uso, de la recomendación por parte del host, o de la adecuación del mismo a las funcionalidades ofrecidas por el propio host. Incluso alguno de ellos facilita con el contrato un software de comunicaciones y terminal especialmente diseñado para la consulta de sus bases de datos.

7.4. El IRS (Information Retrieval System/Sistema de Recuperación de Información).

Una vez conectado el usuario al servicio, y validada su autorización, se encuentra en situación de utilizar el software de interrogación y recuperación que pone a su disposición el host. A este tipo de sistemas informáticos se les engloba bajo el término IRS (Information Retrieval System, Sistema de Recuperación de Información) (MEADOW, 1992). Los IRS son aplicaciones diseñadas ex-profeso para la consulta de bases de datos, por lo que sus funcionalidades se centran en facilitar al usuario la ejecución de consultas, la revisión y la manipulación de los resultados, y la comunicación de los mismos. Carecen, por lo tanto, de las características relacionadas con la creación y administración de bases de datos. Los IRS se encuentran situados en el ordenador remoto o host, en el cual ejecutan

las órdenes remitidas por el usuario, a través del software de comunicaciones y/o terminal.

Las funciones más comunes desempeñadas por los IRS (HAWKINS *et alii*, 1985) corresponden a:

1. Acceso y notificación: funciones relacionadas con la conexión y proceso de autentificación de un usuario, así como notificación de las modificaciones, novedades, etc. del servicio que se le presta.
2. Selección de bases de datos: permite al usuario seleccionar cual de las bases de datos existentes va a consultar. Casi todos los sistemas ofrecen la posibilidad de buscar en varias al mismo tiempo.
3. Búsqueda por fichero inverso o tesauro: el usuario consulta el contenido de estos ficheros para comprobar la existencia de los términos a utilizar en ecuaciones posteriores, las relaciones con otros términos, etc. Este tipo de consulta no supone la recuperación de registros.
4. Creación de subconjuntos: la respuesta a una ecuación crea un subconjunto virtual de documentos, sobre los que es posible ejecutar combinaciones con otros subconjuntos, o acciones de manipulación.
5. Búsqueda de cadenas de caracteres: funciones de búsqueda de texto (analizadas en el capítulo dedicado a las bases de datos documentales).
6. Análisis de subconjuntos: ofrecen funciones que permiten un somero análisis numérico del número de ocurrencias, frecuencias de años de publicación, etc.
7. Ordenación, visualización y formato de registros: funciones relacionadas con la presentación por pantalla de los registros, según criterios de ordenación, uso de diferentes formatos de visualización, etc.
8. Impresión, descarga y solicitud: tanto en el ordenador local como en el host, se incluye la posibilidad de enviar los registros por correo electrónico. Además, es posible, en algunos casos, obtener copias de los registros para uso local, en formato electrónico, mediante teledescarga. Por último, ofrecen la posibilidad de solicitar una copia del documento primario.
9. Archivo, ejecución y edición de ecuaciones: los host ofrecen a sus usuarios la posibilidad de almacenar las ecuaciones de búsqueda de su interés, para ejecutarlas en sesiones posteriores, modificándolas si es preciso. De esta forma pueden incluso crear perfiles de usuario, que faciliten las búsquedas retrospectivas, y la difusión selectiva de la información.
10. Control de sesión: información al usuario de los tiempos de conexión, ecuaciones ejecutadas, documentos obtenidos y los costes derivados de todo ello.

```
                         CORDIS Main Menu Version 3.1
------------------------------------------------------------------------------
             COmmunity Research and Development Information Service

NOTE : Menu Version 3.1 changes in database menus released on 1995-10-05;
       for details, type '88'.

What would you like to see?

 1+  A CORDIS database

 2   Information on CORDIS and on CORDIS databases

 3   Complementary RTD-related information

 5   General instructions for new users

Type option number, then press ENTER
------------------------------------------------------------------------------
 8 Help       9 Exit      88 Menu Version 3.1 Readme File
```

Fig.7.1. Una sesión en el host ECHO.

7.5. La creación de bases de datos en línea.

La teledocumentación es, como se ha apuntado, un negocio. Como tal, la existencia de una base de datos consultable en línea, mediante un servicio comercial, responde a una decisión tomada en tal sentido por parte del productor. El factor decisivo es la existencia de un volumen considerable de usuarios, dispuestos a pagar por acceder a la información que pueda contener la futura base de datos. Una vez tomada la decisión inicial de crear la base de datos, es necesario adoptar otras decisiones complementarias que atañen a los siguientes aspectos:

1. Cobertura de la base de datos: área o campo que va a cubrir la base de datos.
2. Diseño de los registros individuales: determinación de las categorías de datos que permitirán representar a las entidades.
3. Selección de las fuentes primarias: selección de las fuentes y de los documentos de los que se recopilará información para ser introducida en la base de datos, de la cobertura temporal y espacial de los mismos, etc.
4. Creación del contenido de los registros: introducción de los datos pertinentes, tomando en consideración las reglas de descripción y de uso del lenguaje, de catalogación, de composición de consultas...
5. Mecanismos de introducción, manipulación y actualización de los datos.
6. Control de calidad: mecanismos que aseguren la adecuación del contenido de los registros a la realidad, y la satisfacción del usuario.

Puede apreciarse que las cuestiones señaladas resultan similares a las vistas en el capítulo dedicado a los SGBD y a la creación de bases de datos. El

diseño y desarrollo de los recursos de información electrónica, que tienen como soporte principal bases de datos, siguen las normas y principios generales enunciados al tratar las bases de datos.

7.6. Clasificación de las bases de datos.

Los distribuidores pueden ofrecer bases de datos de contenido temático muy variado: medicina, política, derecho, cultura... por esta razón algunos autores han adoptado una clasificación basada en los trabajos de Fundesco (FUNDESCO, 1986) y de otras fuentes, que tienden a clasificar las bases de datos según el siguiente esquema:

1. Bases de datos referenciales: son aquellas que ofrecen registros que su vez son representaciones de documentos primarios. Dentro de éstas distinguen, as su vez, entre
    1.1. bibliográficas: aquellas cuyo contenido son registros de tipo bibliográfico.
    1.2. directorios: aquellas cuyo contenido está referido a la descripción de otros recursos de información, como por ejemplo un directorio de bases de datos.
2. Bases de datos fuente: serían aquellas bases de datos que ofrecen el documento completo, no una representación del mismo. También distinguen entre:
    2.1. numéricas: contienen información de tipo numérico, como un censo o indicadores cuantitativos.
    2.2. textuales: ofrecen el texto completo de un documento.
    2.3. mixtas: combinan ambos tipos de información, como por ejemplo informes económicos de empresas, datos geoeconómicos, etc.

De esta clasificación puede deducirse, evidentemente, que la información obtenida podrá ser referencial o factual. Sin embargo, cabe plantear que las modificaciones que están siendo introducidas en el ámbito de las bases de datos, tanto a escala tecnológica como de contenido, como por ejemplo las bases de datos multimedia, introducirán variaciones en esta clasificación, que deberá adoptar criterios más flexibles.

7.7. Fases tradicionales de una búsqueda teledocumental.

La bibliografía ha señalado como clásicas las siguientes fases en la ejecución de una búsqueda teledocumental (BATES, 1987):

1. Conexión con el host, identificación del usuario y autorización de acceso.
2. Selección del fichero o ficheros de datos a consultar.
3. Estudio de los descriptores a través de los tesauros.
4. Agrupación de los descriptores sinónimos en diferentes grupos.
5. Establecimiento de las relaciones existentes entre los conjuntos.

6. Combinación de los descriptores que representan a los conjuntos.
7. Formulación y ejecución de la ecuación de búsqueda.
8. Visualización de los registros resultantes.
9. Impresión o importación de los registros.

Este esquema tradicional resulta inadecuado en la actualidad (BASCH, 1993). Una búsqueda de este tipo consume gran cantidad de recursos económicos, tanto de comunicaciones como de uso del host, que son facturadas al usuario, lo que repercute en un aumento del precio final de acceso a la información. Más correcta resulta la ejecución de los pasos 2 a 6 en modo local. Es decir, utilizar los recursos facilitados por el distribuidor con el contrato como herramientas locales para agilizar estas tareas, de tal forma que en el momento de conexión al host se sabe positivamente cuales son los ficheros a utiliza, los descriptores pertinentes, y las relaciones entre ellos. El esquema propuesto obvia, además, la casi necesaria revisión de las ecuaciones, por lo que será recomendable efectuar un somero análisis de los documentos resultantes de una búsqueda entre los puntos 7 y 8, antes de ejecutar órdenes que puedan suponer una sobrecarga para el usuario.

Otra cuestión clave a considerar es la presencia de diferentes IRS en los hosts disponibles, lo que significa, a su vez, la necesidad de conocer y manejar varios lenguajes de interrogación (ARMSTRONG y LARGE, 1992). El usuario se ve obligado a conocer las características de los diferentes lenguajes, así como las particularidades de cada sistema, lo que supone un inconveniente para la optimización de la recuperación de información mediante teledocumentación.

7.8. Estructura de los registros existentes en los host.

Los registros que se incluyen en las bases de datos ofrecidas por los host suelen ofrecer una estructura muy similar, especialmente los contenidos en las bases de datos bibliográficas referenciales (FUNDESCO, 1986). Las categorías de datos contenidas responden a:

1. Número de identificación del registro.
2. Título.
3. Autor (en ocasiones se diferencia entre el primer autor y el resto).
4. Centro de trabajo (investigación, etc.) en el que se localiza el autor del trabajo.
5. Cita bibliográfica completa (obra, número, año, páginas, etc.)
6. Resumen.
7. Descriptores o encabezamientos.
8. Identificadores.

De esta forma combinan la necesaria estructuración de la información, con las categorías de datos que permiten describen cada uno de los recursos de información. La utilización de técnicas de manipulación y recuperación de grandes extensiones de texto es similar a la analizada al tratar los sistemas de gestión de bases de datos documentales.

7.9. Servicios complementarios facilitados por los distribuidores.

Los distribuidores o host no suelen limitar sus servicios a la puesta en línea para su consulta de las bases de datos. Junto a esas consultas, ofrecen otros tipos de servicios. Entre estos merecen destacarse los referidos a búsquedas retrospectivas y a las copias de documentos primarios.

Las características de las bases de datos y de los lenguajes de interrogación permiten realizar búsquedas retrospectivas por parte del usuario, facilitar la difusión selectiva de la información, y establecer servicios de alerta sobre los temas de interés. Hay que destacar que los servicios de alerta, junto a la creación de periódicos o boletines personalizados de información, son una de las áreas de mayor expansión e investigación en el mercado de la información estadounidense, que suele marcar las pautas en estas cuestiones.

La obtención de copias del documento original es otro de los servicios complementarios ofrecidos por los distribuidores, que pueden hacerlos llegar al usuario por diferentes vías, con criterios de coste y oportunidad sumamente variados.

Otro de estos servicios es la teledescarga, entendiendo como tal la obtención de una copia del documento o registro, sea referencial o factual, y su almacenamiento en el ordenador local. Es imposible controlar la teledescarga de registros por parte de los usuarios, por lo que los host han optado por aplicar diferentes medidas, que van desde la liberalización de la misma, hasta la carga de cuotas por cada documento que es visualizado por el usuario, y que por lo tanto puede ser susceptible de teledescarga.

7.10. Ventajas e inconvenientes de la teledocumentación. El estudio del coste.

Al igual que todo servicio de acceso a la información, la teledocumentación ofrece ventajas e inconvenientes. Como ventajas deben señalarse el acceso inmediato a gran volumen de información, exhaustiva y muy completa, además de actualizada casi instantáneamente. Se dispone además de conocimiento sobre literatura gris o de difícil acceso. La obtención de los documentos pertinentes puede realizarse a través de varias vías, mediante la modificación inmediata de las ecuaciones de búsqueda, o el replanteamiento de los conceptos utilizados.

Como limitaciones deben señalarse la ausencia, en numerosas ocasiones, de información antigua, anterior a la creación de la base de datos, y la falta de bases de datos en áreas consideradas económicamente poco rentables.

No existe un único esquema de tarifas para la teledocumentación. La aplicación de tarifas puede variar entre los diferentes distribuidores, con motivo de lanzamiento de bases de datos o nuevos servicios, competitividad, etc. Incluso en

ocasiones se realizan "ofertas" de conexión y utilización de bases de datos. Independientemente de esto, hay que tener en cuenta que los hosts suelen cargar en la factura del usuario como conceptos tiempo de conexión, número de ecuaciones ejecutadas, bases de datos utilizadas, número de documentos visualizados, términos consultados en los tesauros, uso de correo electrónico interno o espacio utilizado para el almacenamiento de ecuaciones de búsqueda. Esto supone que el documentalista debe controlar rigurosamente la utilización de los servicios del hosts, de forma que el coste de la búsqueda se adecue lo más posible a los resultados de información obtenidos.

8. INTERNET.

Lecturas recomendadas: HAHN, 1994; GARCÍA y TRAMULLAS, 1996; UBIETO, 1994, 1995; GILSTER, 1996; TRAMULLAS, en prensa, sin publicar; CODINA, 1996; SCOTT BRAND, 1996; ORTIZ REPISO y OLMEDA, 1994; TSENG, POULTER Y HIOM, 1995.

8.1. Introducción

El presente capítulo pretende dar una visión general del nacimiento y desarrollo de Internet, de los principios técnicos sobre los que se sustenta, y de las herramientas disponibles para la explotación de los recursos de información accesibles. Es necesario hacer una precisión. Se aprecia en estos momentos una peligrosa "inflación" sobre Internet. Son excesivas las referencias y trabajos sobre la misma y sus aplicaciones, sin que en realidad, sean abundantes las aportaciones novedosas, lo que da como resultado la creación de un panorama engañoso.

El interés por la comunicación remota entre ordenadores es temprano en la comunidad científica, por el interés que despertaba poder disponer de información y datos actualizados en lugares alejados de los centros de investigación. Sin embargo, las redes existentes hace fechas relativamente recientes no eran adecuadas para este tipo de aplicación, o bien resultaban excesivamente costosas. Los mecanismos clásicos de la teledocumentación, vistos con anterioridad, era los más utilizados para acceder a información actualizada sobre el estado de los conocimientos en cualquier materia.

La expansión de los recursos telemáticos a nivel científico internacional ha paliado, en gran parte, este tipo de problemas. Es posible acceder, en estos momentos, mediante redes como "Internet", a grandes cantidades de información de todo tipo, y mantener contactos regulares y actualizados con grupos de trabajo en toda la extensión del planeta. Por supuesto, esto supone un variado y rico intercambio científico y personal, y, lo que resulta fundamental, muy rápido y

prácticamente gratuito. Como complemento, también se han utilizado, en la medida de los posible, otras redes de comunicaciones de acceso público.

Para el entorno científico y académico español ello es posible gracias a la conexión de un número creciente de instituciones españolas de I+D a la "Internet" mundial. El coste de las líneas telemáticas, fijo e independiente del uso, es asumido por el "Plan Nacional de I+D" materializado en la red ARTIX. A nivel práctico esto supone —por el momento— un coste casi nulo para las instituciones y sus miembros. Para el ciudadano particular se han puesto en marcha numerosas empresas privadas que ofrecen acceso a Internet mediante el pago de una cuota, que varía dependiendo de tiempo de conexión, servicios, etc. A estas empresas se accede actualmente a través de una línea especial de comunicaciones dispuesta por Telefónica, a la que se conoce con el nombre de “Infovía”.

El rápido fluir de la información es un factor determinante en su utilización. La difusión es un elemento imprescindible para su aprovechamiento, y las redes académicas ofrecen el soporte adecuado para ello. Las necesidades del usuario son muy diversas, por lo que se han establecido una serie de mecanismos sobre estas redes que procuran servir las necesidades de información de los investigadores. La búsqueda de recursos informativos se convierte entonces en un navegar interactivo, para el que es necesario conocer y aplicar las herramientas existentes.

8.2. La red Internet.

Actualmente, la Internet es la interconexión de resdes informáticas a escala global y planetaria que engloba a todas las redes que utilizan protocolos TCP/IP (HAHN, 1994). No existe un órgano de gobierno en la red, si bien la Internet Society promueve la utilización de la red y procura coordinar su desarrollo. A ello ayudan el Internet Advisory Board, con recomendaciones técnicas y protocolos, y el Internet Engineering Task Force, comité técnico de voluntarios que ofrecen soluciones técnicas.

A finales de los años 60, una sección del Departamento de Defensa de los Estados Unidos desarrolló una red de comunicaciones, denominada ARPANET. Su finalidad era conectar centros de investigación para el desarrollo de proyectos, y debía ser capaz de soportar ataques y destrucciones sin perjuicio de su capacidad de transmisión. Los protocolos utilizados fueron los que se llamaron TCP/IP (Transfer Control Protocol/Internet Protocol), y fueron incluidos en el UNIX de Berkeley, lo que aseguró su difusión. La evolución de esta red y la aparición de otras similares durante los 70 y primeros 80, hicieron que la National Science Foundation (NSF) se plantease, a mediados de los 80, la unión de las diferentes redes existentes, tanto de la Administración como de centros de investigación y universidades, con nuevos centros de supercomputación que se estaban creando. Esto se realizó utilizando ARPANET y NSFnet. El éxito fue tal que en 1987 la red estaba saturada. Desde 1990, ANS, una empresa sin ánimo de lucro, se encarga de gestionar la red y facilitar sus servicios. La red se financia con dinero público, y

las organizaciones y empresas privadas tienen acceso a través del pago de cuotas por los servicios recibidos.

En Europa, en 1986 se funda RARE, que desarrolla el proyecto COSINE, y pretende agrupar las redes de investigación ya existentes (HEPnet, EUnet, EARN), así como marcar un desarrollo común europeo. Esto ha permitido establecer la red europea IXI ahora llamada EuropaNet. Por supuesto, todo este proceso está muy influenciado por condicionamientos políticos europeos. Como usuarios particulares, destaca el trabajo de EUnet (red de usuarios UNIX europeos).

8.3. Equipo básico y protocolos para el aprovechamiento de Internet.

Los medios materiales necesarios para el acceso a la interconexión de redes informáticas a la que se identifica con el nombre de Internet son de uso común gracias a la progresiva penetración de la informática a todos los niveles: una cuenta en un gran ordenador ("workstation", miniordenador o "mainframe"), o un PC u ordenador personal de otro tipo, conectado al mundo exterior (RedIRIS, ARTIX, Infovía), vía serie (modem y línea telefónica con Infovía u otro proveedor) o vía red (Ethernet o similar) (UBIETO, 1995). Los conocimientos técnicos precisos para preparar y aprovechar una conexión pueden resumirse como sigue:

1. El sistema operativo de un ordenador monousuario o multiusuario, el acceso al mismo, gestión de ficheros y directorios, así como los correspondientes sistemas operativos.
2. El proceso de conexión al ordenador central con los medios y programas anteriormente citados, cuando sea necesario.
3. Las convenciones utilizadas en protocolos de comunicaciones, especialmente el "domain name system" de "TCP/IP" y las aplicaciones de "terminal remoto" (TELNET), "transferencia de ficheros" (FTP) y "correo electrónico" ("ean" en VMS, "elm" en UNIX, Eudora o Pegasus en sistemas PC). Y finalmente lo referente a la "emulación de terminal", siendo suficiente utilizar la que se conoce como "VT-100".

El conjunto de software necesario para establecer la comunicación entre los diferentes ordenadores recibió el nombre de TCP/IP (Transfer Control Protocol/Internet Protocol), y surgió varios años antes que el estándar OSI. Su desarrollo fue promovido por su selección por parte de la Administración norteamericana para su red de defensa, y su inclusión dentro del sistema operativo UNIX como elemento de comunicación entre estaciones de trabajo. Su gran éxito fue debido, además, a incluir cierto número de protocolos de alto nivel muy utilizados por los desarrolladores, entre los que destacaban TELNET (que provee al usuario de un interfaz virtual ASCII, simulando la conexión directa), FTP o File Transfer Protocol (que permite la transferencia de ficheros en formato ASCII o IMAGE, o bien BINARY), SMTP ó Simple Mail Transfer Protocol (transferencia de correo de usuario a usuario a través de la red) y SNMP ó Simple Network Management Protocol (conjunto de reglas que permite gestionar los perifericos que soportan este protocolo a través de uno de ellos).

Aunque TCP/IP no cumple los requisitos del modelo OSI, es paralelo a éste. De hecho, las capas 1 y 2 de OSI corresponden a la capa 1 de TCP/IP, la capa 3 de red a la capa de IP, la capa 4 de transporte a la capa de TCP y UDP, y las capas 5, 6 y 7, de sesión, presentación y aplicación respectivamente, son cubiertas por FTP, SMTP, TELNET, etc.

Dado su origen, el modelo TCP/IP se adecuada a una red de área extensa más que a una red local, ya que trabaja con un control de errores más exhaustivo, lo que produce un descenso en el rendimiento, mientras que el modelo OSI es más adecuado para las aplicaciones en tiempo real. Sin embargo, aunque las políticas, por lo menos las europeas iban por este camino, el éxito de los norteamericanos en el desarrollo de herramientas y aplicaciones para Internet, sobre protocolos TCP/IP, y la mayor facilidad para el desarrollo de aplicaciones que ofrecen éstos han provocado que TCP/IP sea *de facto*, el protocolo actual y futuro de Internet.

8.4. Nombres y direcciones en Internet.

Para el usuario resulta evidente la necesidad de conocer no sólo cómo acceder, sino conocer dónde se encuentran los recursos informativos de sus interés. En Internet, todos y cada uno de los computadores conectados poseen una dirección IP, un conjunto organizado de números que permite diferenciarlo del resto de ordenadores presentes en la red. Las direcciones IP se componen de cuatro grupos de números, que pueden contener de uno a cuatro caracteres, o de o a 255 en valor numérico, separados por un punto, como por ejemplo 155.210.60.60 ó 12.134.145.789.

Dada la dificultad en memorizar y utilizar directamente este tipo de direcciones, se utiliza como sustituto el sistema DNS (Domain Name Server), que sustituye las direcciones IP por direcciones alfanuméricas más sencilla (aunque en algunos casos esto es discutible). Por ejemplo, 155.210.60.60 corresponde a jabato.unizar.es, de más fácil comprensión y lectura. Según las reglas de DNS, la lectura se realiza desde la derecha. El primer grupo (es) corresponde al dominio de primer nivel, generalmente el país; el segundo grupo (unizar) corresponde al dominio de red, mientras que el último es el nombre del ordenador (jabato). Puede aparecer un intermedio entre el nombre de la máquina y el de la red, que puede corresponder a una red local o subdominio. Los dominios de primer nivel corresponden a la norma ISO identificativa de paises, con la excepción de los dominios .mil (militar), .gov (gubernamental), .edu (educativo), .com (comercial) u .org (ONG), que corresponden a entidades situadas en los Estados Unidos de América.

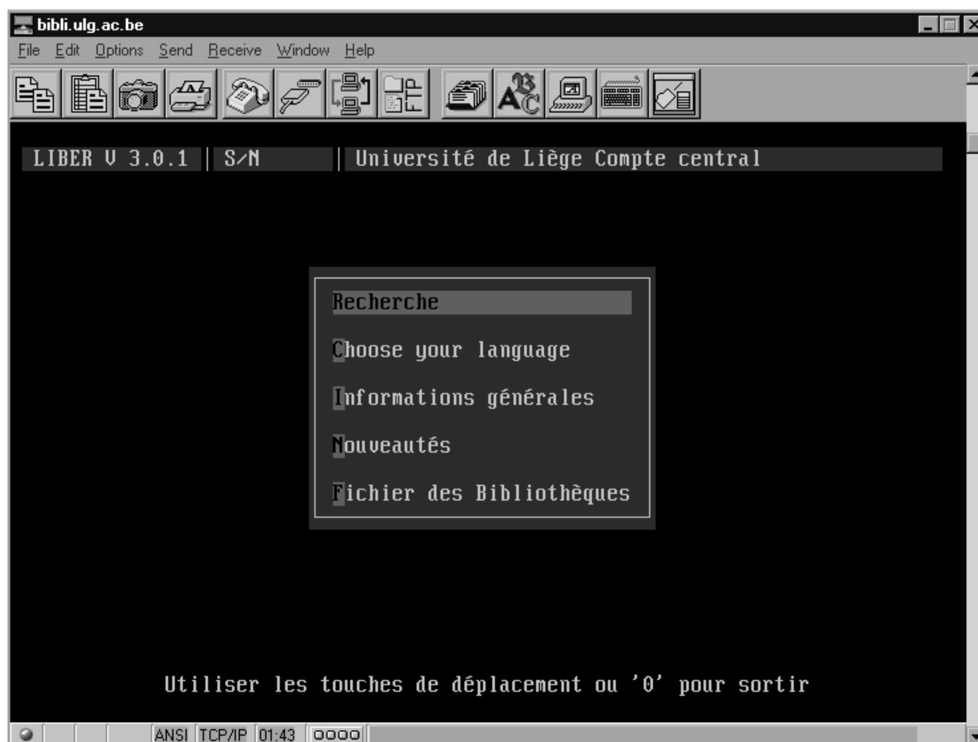


Fig.8.1. Conexión Telnet a la Biblioteca de la Universidad de Liège.

8.5. Aplicaciones de búsqueda y recuperación de información sobre Internet.

Sobre el protocolo TCP/IP se han establecido, utilizando las facilidades y fiabilidad que éste aporta, un conjunto de aplicaciones especializadas para facilitar el acceso y la transferencia de información de forma fiable entre los usuarios, independientemente del sistema operativo sobre el que trabajen (UBIETO, 1994; HAHN, 1994). De esta forma, pueden encontrarse actualemnte gran cantidad de "clientes", es decir, aplicaciones que permiten interactuar con "servidores" de la misma clase, normalmente gratuitos, y disponibles para las principales plataformas de ordenadores personlaes, con sistemas operativos e interfaces DOS, UNIX y Macintosh, y X-Windows y Windows, respectivamente. Entre éstos, deben citarse obligatoriamente los siguientes:

8.5.1. Mensajería:

El correo electrónico (o e-mail) permite el intercambio de mensajes entre personas a lo largo de toda la red. Basta con conocer su dirección electrónica para entrar en contacto con el destinatario, en cualquier punto. Junto al correo personal, deben citarse las listas de distribución, sobre un tema concreto. Los interesados en ese tema se inscriben en una lista de

correo, que actua como mensajero virtual, redireccionando a todos los inscritos los mensajes que se envíen al servidor central.

8.5.2. Noticias:

Denominadas NEWS, completa al servicio de mensajería. Utilizando un símil, sería un tablón de anuncios en el que cualquier usuario deja un mensaje para todos aquellos que deseen leerlo, con la ventaja de que no envía todos los mensajes al usuario, sólo aquellos que desea leer. A los diferentes tablones se les llama grupos. Como las listas de correo, pueden ser públicos o privados. Las news son uno de los servicios de mayor crecimiento y uso dentro de Internet.

8.5.3. Transferencia de ficheros:

Dentro de los computadores conectados a Internet puede encontrarse gran cantidad de información, almacenada en ficheros de base de datos, textos, gráficos, así como software de todo tipo (freeware, shareware, demos...). Mucha de esta información es accesible al usuario, que puede cargar en su propio ordenador los ficheros originales. La aplicación utilizada para estas transacciones es FTP.

8.5.4. Terminal remoto:

Servicio primario, su utilidad es acceder a las cuentas en los ordenadores propios desde cualquier punto de la red, y realizar los trabajos necesarios en cada momento.

8.5.5. Búsqueda de información:

El gran volumen de información existente hace necesaria la utilización de herramientas que permitan su localización (GILSTER, 1996). Dentro de ellas, destacan:

- *WHOIS*: búsqueda de personas. Al introducir el nombre de un individuo, ofrece sus datos profesionales y su localización. Una de sus mayores limitaciones es la falta de coordinación entre países.
- *X.500*: evolución del anterior, ofrece la posibilidad de coordinar los datos locales con los de la red.
- *NETFIND*: búsqueda de personas, partiendo de un conjunto de términos (nombre, profesión, dominio...) que permiten identificarlo. Busca por toda la red y ofrece una lista de posibles candidatos.
- *KNOWBOTS*: servicio de búsqueda de tipo directorio, distribuye consultas por la red hasta localizar el objetivo.
- *ARCHIE*: búsqueda de ficheros. Es una gigantesca base de datos con todos los directorios de ficheros existentes en los ordenadores conectados a Internet.En general, existen servidores Archie en cada país, que incluyen en su ámbito el propio país, áreas cercanas y otros servidores seleccionados de todo el mundo. Admite consulta por patrones (nombres, etc.), y ofrece un listado de todos los ficheros que se acomodan al patrón, así como el computador en el que se encuentran.
- *GOPHER*: búsqueda a través de menús por toda la red. Mantiene un conjunto de ítems de contenido y de otro de ítems de

estructura. Permite el acceso a otros Gophers, la conexión a otros computadores, la transferencia de ficheros, la consulta a bases de datos, el acceso a directorios, etc. Su sistema de menús es programable por el usuario, por lo que éste puede organizar su utilización de los recursos de la red. El espacio de información organizado y accesible de esta manera es llamado "gopherespacio". GOPHER ofrece una herramienta de funcionalidad parecida a ARCHIE, pero dedicada principalmente a documentos incluidos en servidores Gopher, denominada VERONICA.

- *PROSPERO*: esta herramienta permite organizar espacios virtuales con ficheros a lo largo de toda la Internet, mediante enlaces remotos, para cada usuario.
- *WWW (World Wide Web)*: sistema de hipertexto e hipermedia desarrollado por el CERN de Ginebra, organiza la información a lo largo de toda la red mundial según los criterios del creador. Las aplicaciones WWW más extendidas son Netscape Navigator, Internet Explorer y NCSA Mosaic, que ofrecen versiones para varios sistemas operativos.
- *WAIS (Wide Area Information Server)*: herramienta que permite crear índices de los términos contenidos en un documento, y realizar búsquedas sobre esos índices para identificar el documento que interesa al usuario a lo largo de la red. Los servidores WAIS indizan correo electrónico, news, documentos gopher... Utiliza el protocolo Z39.50 de intercambio de información entre bibliotecas.

Las tres últimas herramientas, GOPHER, WWW y WAIS se encuentran en muchas ocasiones integradas, lo que proporciona un poderoso entorno de acceso y recuperación de información.

Otra herramienta que hay que destacar es Hytelnet. Hytelnet (ORTIZ REPISO y OLMEDA, 1994) es una aplicación de tipo hipertextual, desarrollada por P. Scott, que ofrece un exhaustivo conjunto de catálogos de bibliotecas, bases de datos, servicios WAIS, etc, a los que se accede mediante terminal remoto, liberando al usuario de las tediosas conexiones individuales, con lo que se agiliza la consulta de fondos bibliográficos a nivel mundial. Al usuario le basta con seleccionar una de las direcciones presentes para que el sistema establezca la conexión y el acceso oportuno al sistema. Originalmente pensada para utilizar con servicios de terminal remoto como Telnet, en la actualidad existe una versión web, llamada WebCats, que integra el acceso mediante World Wide Web de aquellas bibliotecas en la que esto ya es posible.

Debe señalarse, para finalizar, que muchos de los servidores ftp, gopher y wais pueden encontrarse interrelacionados unos con otros, y, por consiguiente, es posible acceder a la misma base de datos o al mismo fichero de texto a través de tres o cuatro servidores de información diferentes. Por desgracia, esta proliferación de referencias cruzadas no ha sido todavía controlada por ninguna entidad, aspecto éste que se espera sea tratado en breve por el comité correspondiente de la Internet Society.

8.6. El estandarte de Internet: World Wide Web.

La aplicación que más ha favorecido el conocimiento y despegue de Internet en todos los ámbitos ha sido el Worl Wide Web, o telaraña mundial, de tal forma que incluso en ocasiones se llega a confundir la propia Internet con el Web, como también se conoce (GARCÍA MARCO y TRAMULLAS SAZ, 1996). El World Wide Web fue un desarrollo de Tim Berners-Lee, del CERN de Ginebra, que buscaba desarrollar un medio de publicación e intercambio de información entre los físicos de partículas. Sus desarrollos de hicieron públicos en Internet, y esta novedosa forma de organizar y acceder a la información tuvo tal éxito que en estos momentos existe un proyecto internacional, dirigido desde el MIT, y en el que colaboran CERN e INRIA, para lograr un desarrollo adecuado del ámbito Web, y coordinar en lo posible las nuevas aplicaciones y herramientas que aparecen rápidamente.

El principio general que guía el ámbito web es el hipertexto y el hipermedia (véase el capítulo correspondiente). En breve, un autor crea un documento web que puede contener texto, gráficos estáticos y/o dinámicos, sonido...y lo pone a disposición de todos los posibles usuarios. A su vez, el documento puede contener enlaces o "links" a otros documentos web, situados en los más remotos confines de Internet, que pueden tratar de temas relacionados, y cuya consulta puede ser provechosa para el lector. Estos documentos son accesibles a través del protocolo HTTP (HyperText Transfer Protocol). El lector/usuario del documento tiene, a partir de ese momento, la libertad de elegir que camino, que nuevo documento, quiere acceder y consultar. Como puede deducirse, se trata de una manera radicalmente nueva tanto de publicar información, como de acceder a ella, muy diferente al concepto tradicional de teledocumentación, aunque íntimamente relacionado en su objetivo y finalidad.

8.6.1. El cliente web

La herramienta o aplicación utilizada para explorar y navegar por el ámbito Web se conoce como browser o cliente web. Se trata de un cliente capaz de acceder, recibir, montar y mostrar los documentos web. Hay que recordar que también en este caso se está tratando con una arquitectura cliente/servidor, lo que supone la existencia de un servidor web que recibe una petición y envía una respuesta, y un cliente que envia la petición, recibe la respuesta y la muestra al usuario. El cliente web recibe en respuesta a su requerimiento un conjunto de ficheros (texto, gráficos...), y las instrucciones para organizarlos y mostrarlos al usuario tal y como desea el creador del documento. Conforme va recibiendo los componentes, el cliente web los organiza y los va mostrando progresivamente al usuario.

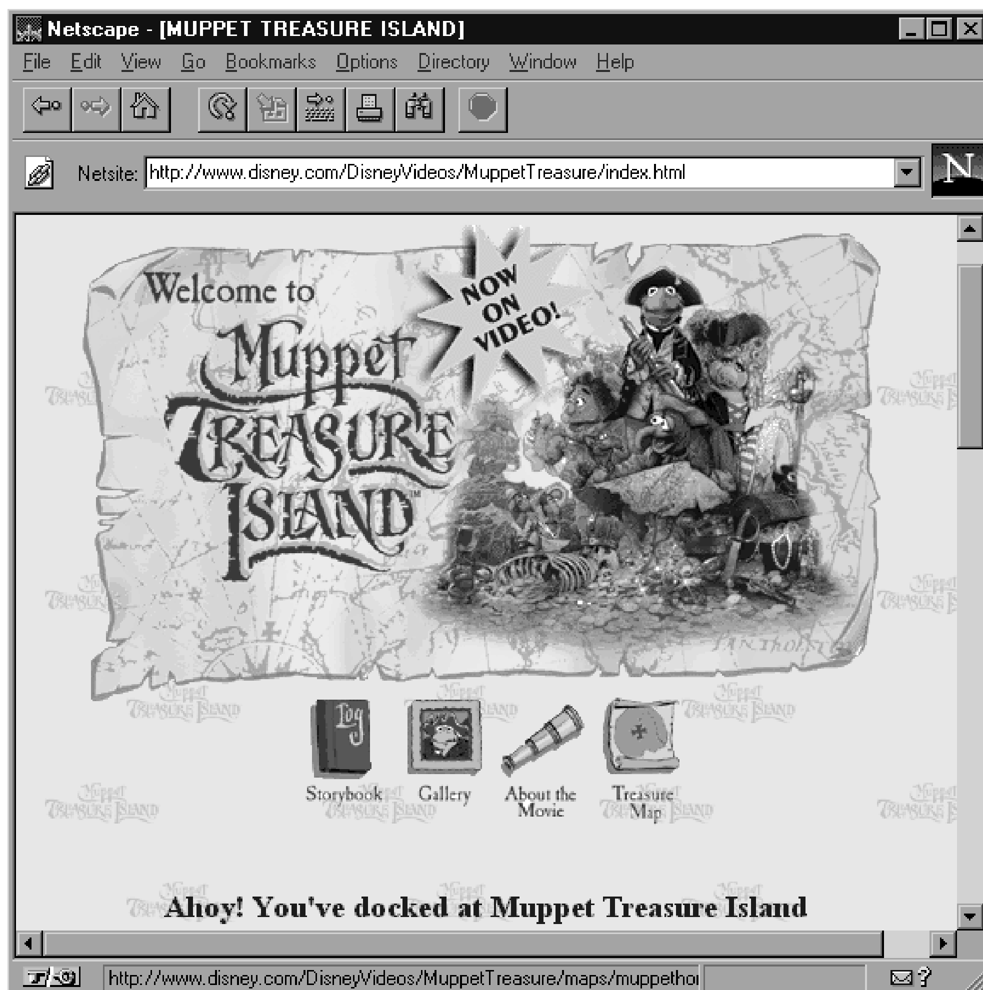


Fig.8.2. Aspecto del cliente web Netscape Navigator.

Las aplicaciones clientes resultan ser en estos momentos sumamente avanzadas, y sus funcionalidades no se limitan a la obtención y proceso de montaje de las páginas web. Son capaces de almacenar en el ordenador local las páginas, las imágenes, sonidos u otro tipo de ficheros que contengan, y mostrarlos posteriormente. Además, ofrecen la posibilidad de establecer “bookmarks” o marcas de lectura, es decir, crear una agenda o listado de páginas de interés para el usuario, accesibles en cualquier momento, sin necesidad de repetir un proceso de navegación para llegar hasta ellas. Incluyen la posibilidad de desarrollar sesiones y acceder a recursos de servidores FTP y Gopher, e incluso a servidores de news, y son capaces de gestionar el correo electrónico del usuario, por lo que se están configurando como una herramienta plenamente integrada. Si a ésto se una la existencia, en cada vez mayor número, de pasarelas web para la consulta

de servidores WAIS, de bases de datos y de catálogos en línea de bibliotecas, puede apreciarse cómo los clientes web se están convirtiendo en la aplicación más universal para la navegación en Internet.

Cabe ahora indicar qué utiliza el cliente web para mostrar las páginas web. El protocolo HTTP utiliza una notación especial para localizar y acceder a las páginas web. Se trata del URL (Universal Resource Locator), una especialización de las direcciones vistas más arriba. Esta notación, orientada, como puede imaginarse, al documento, es un desarrollo del DNS, en el sentido que incluye la dirección de la máquina, pero también la estructura de directorios a la que se debe acceder para obtener el fichero, así como el nombre del mismo. Por ejemplo,

http://jabato.unizar.es/docs/info/saludo.html

es un URL que le está diciendo al cliente web que debe utilizar el protocolo http para comunicarse con la máquina jabato.unizar.es, acceder a la estructura de directorios /docs/info/, y dentro de este último obtener el fichero saludo.html. Todos los URL siguen el mismo esquema, por lo que su utilización resulta intutitiva y sencilla.

8.6.2. Recuperación de información en el ámbito Web.

La proliferación, en términos exponenciales, de documentos web con todo tipo de contenidos informativos, hace necesaria la creación de sistemas organización y recuperación que faciliten el acceso del usuario a aquellas páginas web que le sean de utilidad (GARCÍA MARCO y TRAMULLAS, 1996). En Internet se han adoptado dos enfoques principales:

1. La creación de listados, índices y catálogos ordenados por áreas o materias, de forma que el usuario disponga de un conjunto de fuentes seleccionadas en las que empezar a buscar. El ejemplo más conocido es el norteamericanoYahoo!. Estos catálogos se suelen compilar y organizar de forma casi manual.
2. La creación de índices mediante unas aplicaciones que rastrean o exploran todo el ámbito Web, llamados robots, spiders o wanderers. Estos sistemas, aunque automáticos, ofrecen limitaciones en cuanto a cobertura, nivel de indización del documento u otras cuestiones.

La búsqueda mediante catálogos no suele ofrecer mayor dificultad que la de navegar, de forma guiada, por la estructura jerárquica hasta encontrar la página que contiene las referencias de interés. En el caso de utilizar las bases de datos creadas por un robot (robot que explora Internet localizando e indizando el contenido informativo de los servidores web), la búsqueda se ejecuta, en general, en un proceso similar a las fases clásicas vistas en la teledocumentación, y con unos criterios igual a los vistos en la recuperación de información. Es decir, que hay que componer expresiones o ecuaciones de búsqueda en la que se combinan los términos que describen el área de interés del usuario mediante operadores booleanos, de posición, truncamientos... (CODINA, 1996). La respuesta consiste en una página web que incorpora una somera descripción de las páginas que más se ajustan a la expresión de búsqueda y, lo que es más importante, un enlace

directo a las mismas, mediante el cual el usuario puede acceder al documento original en ese mismo instante.

Un desarrollo más avanzado de la recuperación de información consiste en la utilización de agentes personales (TRAMULLAS, en prensa). Se trata de aplicaciones que se instalan en el ordenador del usuario, y que son capaces de enviar la misma expresión de búsqueda a varios motores de búsqueda al mismo tiempo. La respuesta recibida es filtrada según diferentes criterios (eliminación de duplicados, ponderación de los documentos, creación de índices...), y preparada para su presentación al usuario, Todas estas tareas se desarrollan en un segundo plano, liberando tiempo de trabajo del usuario para otras tareas.

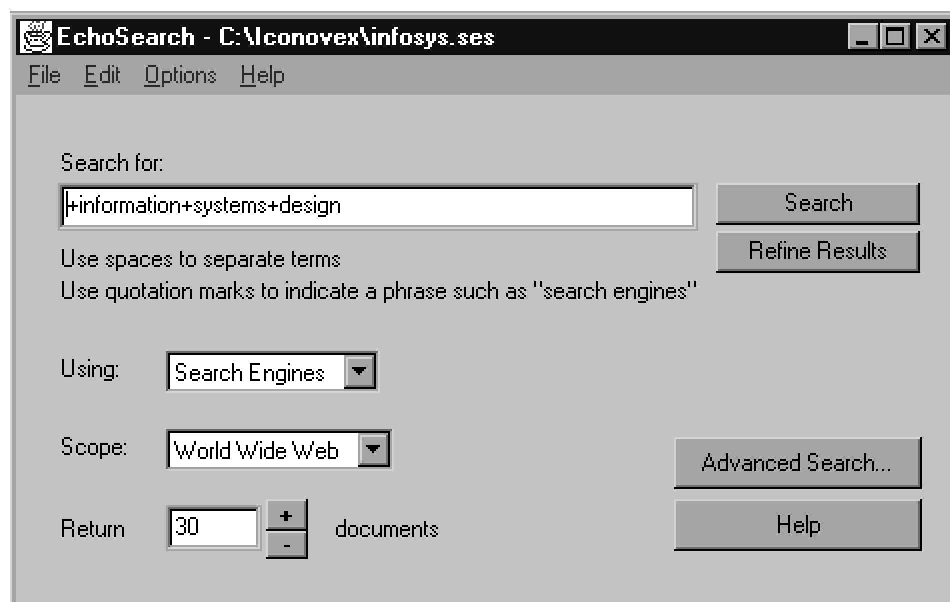


Fig.8.3. EchoSearch: un agente personal

Sin embargo, la recuperación de información en Internet exige del usuario la combinación del enfoque tradicional de interrogación de bases de datos, en esta caso los citados robots, con la navegación derivada de la consulta de las respuestas. Se trata, entonces, de una combinación de “search” y de “browsing”, si se prefieren los términos anglosajones. El usuario, por tanto, no debe limitarse pasivamente a recibir una respuesta a su requerimiento: debe trabajar activamente con esa respuesta, separar lo pertinente del ruido, y utilizar la respuesta como un punto de partida (SCOTT BRAND, 1996). Como se deduce, la recuperación de información en Internet es dinámica y activa, y resulta necesario un conocimiento básico del tema objetivo de la búsqueda, ya que ésta no puede limitarse a utilizar como criterio la aparición de un conjunto de términos, sino que resulta necesario acceder al documento original y validar su contenido, sin olvidar que a su vez puede ser origen de una navegación. La respuesta a una consulta no

se limita a un conjunto de documentos; también puede dar como resultado documentos que son, a su vez, recopilaciones de documentos útiles y pertinentes, y que pueden ser ofrecidos al usuario como un producto documental útil.

Una breve reflexión sobre el tema lleva a considerar que, en realidad, el ámbito World Wide Web puede ser visto como una gigantesca estructura documental distribuida, en la que los documentos y páginas web corresponden a los registros o documentos clásicos desde el enfoque de una base de datos documental. Aún más, resulta que esos documentos responden a muchas de las características que ofrecen los documentos hipermedia construidos gracias a bases de datos multimedia y repositorios de información, como se ha visto en el capítulo pertinente. Y por último, la recuperación de información se realiza a través de la formulación de expresiones o ecuaciones de búsqueda, o a la utilización de índices y catálogos temáticos, en la más pura tradición de la "*Information Retrieval*". Se puede establecer que el usuario del ámbito Web encuentra combinados en éste las ideas de Otlet, Bush, Nelson... junto a las más modernas tendencias teóricas de las ciencias de la documentación y la información y la aplicación práctica de las mismas en la publicación electrónica y la búsqueda y recuperación de información.

8.7.El concepto de Intranet.

De forma muy resumida, Intranet es el término que describe la implantación de las tecnologías de Internet dentro de una organización, más para su utilización interna que para la conexión externa (TRAMULLAS, sin publicar). Esto se realiza de forma que resulte completamente transparente para el usuario, pudiendo éste acceder, de forma individual, a todo el conjunto de recursos informativos de la organización, con un mínimo coste, tiempo y esfuerzo. Intranet e Internet, desde nuestro punto de vista, son casi por completo distinciones semánticas, más que tecnológicas.

Intranet utiliza exclusivamente el modelo World Wide Web, adaptado a su situación y estructura interna, de forma que esta información quede en los límites planteados por la propia organización. Los miembros de la misma utilizarán, como es presumible, clientes web para acceder a la información. Se implantarán, por lo tanto, protocolos TCP/IP, y se utilizará el HTML para la creación de documentos. La disparidad de plataformas y sistemas informáticos existentes en una organización, y los problemas para compartir información entre ellos, fuerzan a los responsables de los sistemas de información a buscar soluciones de integración, de resultados fiables y de un coste aceptable. La utilización de la tecnología World Wide Web, por su facilidad de implantación, su bajo coste, y la rápida aprehensión y aceptación por parte del usuario, así como por su portabilidad a las diferentes plataformas, y su capacidad para interactuar con aplicaciones diversas mediante la utilización del CGI (Common Gateway Interface), se presenta como especialmente interesante.

Los factores que están influyendo poderosamente en el despegue de Intranet pueden resumirse como sigue:

1. Coste asequible.
2. Fácil adaptación y configuración a la infraestructura tecnológica de la organización, así como gestión y manipulación.
3. Adaptación a las necesidades de diferentes niveles: empresa, departamento, área de negocio...
4. Sencilla integración de multimedia.
5. Disponible en todas las plataformas informáticas.
6. Posibilidad de integración con las bases de datos internas de la organización.
7. Rápida formación del personal.
8. Acceso a la Internet, tanto al exterior, como al interior, por parte de usuarios registrados con control de acceso.
9. Utilización de estándares públicos y abiertos, independientes de empresas externas, como pueda ser TCP/IP o HTML.

El oponente de Intranet en la organización, más que los sistemas operativos de red, con los que puede convivir, por el momento, son las aplicaciones identificadas bajo el concepto *groupware*. Es notoria la comparación, en numerosos estudios, con las funcionalidades ofrecidas por Lotus Notes, tanto a nivel de capacidad de acceso y control de la información, como de coste económico de ambas soluciones (formación, licencias, creación de recursos de información...). De hecho, Lotus Notes ha puesto recientemente en el mercado una solución que integra sus características con servidores web, integrándose dentro de una Intranet. Sin embargo, no debe olvidarse que hay muchas cosas que un servidor web no puede hacer, frente a una aplicación como Lotus Notes (por ejemplo, en lo que respecta a interrogaciones complejas sobre varias bases de datos). Pronto se asistirá a la aparición de nuevas aplicaciones que extiendan las capacidades de Intranet hacia el groupware.

8.8. La formación de usuarios.

La correcta gestión del acceso y de los recursos de información disponibles en la Internet obliga a pensar en una adecuación del personal responsable de ello y de los usuarios finales. En primer lugar, el equipo técnico debe estar especializado en protocolos TCP/IP, lo cual no plantea problema si la organización ha seleccionado como plataforma el sistema operativo UNIX en alguna de sus versiones. El resto de sistemas operativos presentes en el mercado, como System 7 de Macintosh (o el futuro Copland), Windows 3.x, Windows 95 ó Windows NT, ofrecen una simple integración de los protocolos y aplicaciones TCP/IP, sin complicaciones, e incluso en ocasiones vienen preinstalados.

Resuelto este aspecto, otro problema lo plantea la selección, utilización y formación en los recursos de información. En el ámbito anglosajón está cobrando cada vez mayor importancia en la organización el Information Manager, como persona responsable de la gestión de la información, en su más amplio espectro, dentro de la empresa. Diferentes estudios permiten diferenciar, según su ámbito de trabajo, entre Information Manager, Information Counselor, Information Worker, etc. Sin entrar en tales disquisiciones, si es evidente que el acceso a información

externa no puede gestionarse de la misma forma que la interna, ya que la externa tiene un alto contenido documental que hace necesario un profesional con un perfil bastante definido, obligado a conocer las fuentes de información externas de interés para el proyecto o investigación en curso.

Por todo lo anterior, el perfil deseable para el responsable de esta gestión debería reunir en sí experiencia y conocimientos de gestión de la investigación, de documentación, de fuentes de información, de técnicas de comunicación, de tecnologías de la información y de sistemas de información, entendiendo este último como sistemas completos, con los recursos humanos, informativos y tecnológicos consiguientes, más allá del concepto tradicional de sistema de información.

8.9. La universalización de los sistemas de información.

La disponibilidad de las redes de telecomunicaciones y de las aplicaciones de todo tipo que pueden ejecutarse sobre ellas hacen posible aumentar la interacción continua entre los diferentes grupos de investigadores. De la misma forma, cambian los conceptos tradicionales de publicación y de acceso a la información, modificando los patrones mentales tradicionales.

En primer lugar, debe señalarse la posibilidad de acceder a repositorios de datos de forma casi instantánea, lo que facilita la comprobación de hipótesis y la mejora de los datos básicos para el desarrollo de proyectos de investigación. Los medios actuales hacen muy fáciles estas tareas, superando los problemas existentes, y actúan como factor de calidad en el proceso investigador. En segundo lugar, se modifica la relación existente entre los propios investigadores. La comunicación entre ellos, de forma dinámica y rápida, facilita el intercambio de experiencias, y actúa como factor unificador en lo referente a terminología, métodos, etc. Obliga a establecer unos estándares, *de facto*, para muchas cuestiones, y establece la publicación electrónica como una forma más de difusión científica.

Las autopistas de la información, junto a la cada vez más cercana generalización de las bases de conocimiento, son el nuevo ámbito de investigación, desarrollo y servicios de los profesionales de la información y la documentación. Las últimas tendencias de los distribuidores comerciales de bases de datos, clásicos recursos accesibles por teledocumentación, indican la progresiva migración de estos servicios al entorno World Wide Web, lo que modificará ineludiblemente la orientación y el trabajo de los expertos en información y documentación, cuyo ámbito de trabajo natural se encontrará en estos sistemas.

9. APROXIMACIÓN A LOS SISTEMAS DE INFORMACIÓN

Lecturas recomendadas: WILSON, 1993:;BURCH y GRUDNITSKY, 1992; CODINA, 1994a, 1994b; SENN, 1992; LUCAS, 1987; SEGIET y BROUSSE, 1992; VICKERY y VICKERY, 1992; ROWLEY, 1994; MIGUEL y PIATTINI, 1993.

9.1. Concepto de sistema.

Un sistema (WILSON, 1993) es un conjunto de elementos interrelacionados en función de un objetivo común, actuando en determinado entorno y con capacidad de autocontrol. De esta definición se deduce que los sistemas se compone de objetos, los cuales tienen unas determinadas características o atributos, sirviendo además los objetos que el forman el sistema para delinear sus límites o fronteras. Estos elementos se unen mediante relaciones, siendo las relaciones entre sistemas llamadas, en ocasiones, interfaces. La principal división que se hace en los sistemas es entre abiertos y cerrados. Un sistema será abierto si intercambia información, energía o elementos físicos con su entorno, y cerrado si no lo hace.

Los principios en los que se basan los sistemas son:

1. Subsidiariedad: todo sistema es subsidiario de otros que lo rodean e interactúan.
2. Interacción: las acciones desarrolladas por un sistema influyen en el comportamiento de los demás.
3. Determinismo: todo hecho que sucede en los sistemas tiene causas definidas y contrastables.
4. Equifinalidad: el diseño del sistema le debe permitir alcanzar su objetivo por medios y acciones diferentes.

Las características que definen a los sistemas son:

1. Estabilidad: cualidad por la que el sistema funciona eficazmente.
2. Adaptabilidad: cualidad que le permite evolucionar dinámicamente con su entorno.
3. Eficiencia: cualidad que permite al sistema alcanzar su objetivo con economía de medios.
4. Sinergia: la capacidad de actuación del sistema es superior a la de sus componentes individuales.

Una vez establecidas las características que definen a los sistemas, éstos pueden diferenciarse en diferentes clases, que pueden encontrarse asociadas entre ellas:

1. Sistemas naturales
2. Sistemas físicos diseñados.
3. Sistemas abstractos diseñados.
4. Sistemas de actividad humana.
5. Sistemas trascendentales

9.2. Sistema e información.

El concepto básico relacionado con un sistema, especialmente en aquellos que se definen como diseñados, es el de proceso de transformación. Las acciones contenidas en el sistema son el conjunto de acciones interrelacionadas necesarias para transformar las entradas en salidas. Esto implica que debe existir un grado de interconectividad entre las entidades que forman el sistema, y permite definir los recursos del sistema. La acción, la transformación que realiza el sistema debe tener disponibles mecanismos que aseguren su control para alcanzar la mejor realización posible, lo cual se consigue mediante medidas de desempeño, utilizadas por procedimientos de toma de decisiones para ejecutar mecanismos de control. El elemento común a todas estas acciones es la información, que se dirigen entre sí las partes del sistema, mediante mensajes, para asegurarse de que se cumplen los controles, y que por lo tanto el sistema cumple su misión adecuadamente. Estas acciones tienen lugar en un área definida, lo que permite establecer los límites del sistema. Por lo tanto, el sistema utilizará, en primer lugar, una información interna, sobre sí mismo, que adecuará su acción a su entorno. Esta información será independiente de su objeto de trabajo. Por supuesto, si se trata de un sistema de información, ésta ofrecerá dos niveles, correspondientes a la propia del sistema, por una parte, y a la información gestionada, por otra.

Desde la perspectiva de la Ciencia de la Información y la Documentación, hay que volver a los principios defendidos por Borko, que sirven para enmarcar el objeto de investigación dentro de la disciplinas de la Ciencia de la Información y la Documentación:

> "Ciencia interdisciplinaria que investiga las propiedades y comportamiento de la información, las fuerzas que gobiernan el flujo y uso de la información y las técnicas manuales y mecánicas

> del proceso informativo para el más eficaz almacenamiento, recuperación y diseminación.”[1]

En la propia definición puede encontrarse un paralelismo muy marcado en los elementos y en la finalidad del sistema de información. Cada vez en mayor medida se está desarrollando un corpus teórico sobre la organización y tareas del documentalista dentro de este ambiente dinámico de los sistemas de información, por lo que se puede afirmar que la gestión de la información, tal y como se acepta en la actualidad, es la integración, en un nuevo marco, del trabajo del documentalista y las posibilidades ofrecidas por los sistemas de información (VICKERY y VICKERY, 1992: 210).

> “The practical work of information scientifist is concerned with development, operation, and management of information systems of all kinds, and with all the various techniques and technologies by which information is transferred.”

Dentro de los diferentes tipos de sistemas, es posible situar a la Documentación como sistema abierto, en relación con las otras disciplinas con las que se relaciona, así como con su finalidad intrínseca. Se trata, entonces, de un sistema que intercambia energía con su ambiente, según diferentes escalas. Las leyes básicas del sistema documental, tal y como fueron establecidas por Ranganathan (RANGANATHAN, 1993), son un elemento de gran valor en el momento de considerar un sistema de información basado en ordenadores:

1. Los documentos son para usuarios.
2. A cada usuario su documento.
3. A cada documento su usuario.
4. Salvaguardar el tiempo del usuario y del documentalista.
5. El servicio documental siempre está en crecimiento.

9.3. Hacia una definición de los sistemas de información.

Toda organización tiene un sistema de información, independientemente de que sea automatizado o no. En cierto modo, se trata de una entidad abstracta, un medio pr el cual los datos y la información fluyen de una persona o departamento a otros, y que puede ser cualquier cosa, desde una comunicación verbal a complejos sistemas de cómputo. Por esta razón, se trata de un sistema abierto, que interacciona con su medio ambiente, con un importante mecanismo de control. Debe considerarse que la organización, ineludiblemente, mantiene flujos de información con el exterior, identificados como externos, y en su propio interior, o internos, los cuales a su vez pueden ser ascendentes, descendentes u horizontales, y tanto de tipo formal como informal. Tomando en consideración las

---

1 BORKO, H. “Information Science: What it is?” *American Documentation*, 19, 1, 1968, p. 145 (citado por LÓPEZ-YEPES, J., “Introducción al estudio de los sistemas de información y documentación en las organizaciones.” en LÓPEZ-YEPES, J. (comp.) *Fundamentos de Información y Documentación*. Madrid: Eudema, 1989, p. 204).

características de los sistemas y las aproximaciones de otros investigadores, y Miguel y Piattini proponen la siguiente aproximación (DE MIGUEL Y PIATTINI, 1993: 17):

> "… definimos el sistema de información como un conjunto de elementos ordenadamente relacionados entre sí de acuerdo con unas ciertas reglas que aporta al sistema objeto (es decir, a la organización a la cual sirve y que le marca las directrices de funcionamiento) la información necesaria para el cumplimiento de sus fines, para lo cual tendrá que recoger, procesar y alamacenar datos, procedentes tanto de la misma organización como de fuentes externas facilitando la recuperación, elaboración y presentación de los mismos."

El sistema de información tienen la misión de memorizar y de producir informaciones y ponerlas a disposición de la organización. Cualquier tipo de sistema de información se crea para adquirir, almacenar, organizar y recuperar información, "… conjunto de procedimientos organizados que cuando se ejecutan, proporcionan información para la toma de decisiones y/o el control de la organización. " (LUCAS, 1987: 16). Se debe completar la definición anterior con las palabras de Wilson (WILSON, 224):

> "Los sistemas de información son en realidad sistemas de datos procesados. Sólo se vuelven sistemas de información cuando alguien usa el resultado. Por tanto, un sistema de información debe incluir al usuario…Un usuario recibe información y, en virtud de desempeñar un tipo de actividad (con base en esta información de apoyo), se produce una salida de datos".

9.4. Los sistemas de información documental (SID).

Ha sido Codina quien más recientemente ha intentado conceptualizar los sistemas de información documental, por oposición a los sistemas de información "tradicionales". Para este investigador, los elementos que definen un SID son los siguientes (CODINA, 1994b: 442):

1. Un conjunto de conceptos fundamentales y, en su caso, unas proposiciones asociadas.
2. Un modelo general de un sistema de información documental.
3. Un modelo general de los procesos de transformación que tienen lugar en el interior del sistema.
4. Un proposición general sobre la Documentación y dos proposiciones particulares, relativas al estatuto de la Documentación como disciplina científica y su relación con otras disciplinas cuyo objeto de estudio también es la información en alguna de sus facetas.

Se puede afirmar, sin duda alguna, que los sistemas de información documentales son un tipo de sistemas de información, con unas características particulares, entre las que destacan las funciones de representación y de comparación. Sin embargo, se puede objetar la afirmación de Codina de que "los SID (sistemas de información documental) no modelan sistemas objeto que desarrollan actividades, sino sistemas objeto que constituyen depósitos interesantes de conocimiento". Precisamente, los modernos sistemas de gestión documental almacenan y desarrollan, o por lo menos deberían, actividades relacionadas con los objetivos de la organización, y no pueden, ni deben, limitarse a ser un simple repositorio de documentos, ya que esto los convierte en simples y obsoletos sistemas de bases de datos. Es difícil aceptar la afirmación de su naturaleza probabilística, que eso supone confundir el funcionamiento interno del sistema con el marco cognitivo externo del usuario, aspecto éste que resulta desconocido para cualquier tipo de sistema de información.

9.4.1. La complejidad multifacetada de los sistemas de información.

La complejidad que preside el estudio de los sistemas de información pueden resumirse utilizando las metáforas presentadas por Van Gigch y Le Moigne (VAN GIGCH y LE MOIGNE, 1989). El paradigma que guía el estudio de sistemas de información debe responder a fuentes de conocimiento, objeto de estudio, escuelas de pensamiento, propósitos, metodologías para las actividades y anomalías. La aplicación de la trilogía formada por paradigma, escuela de pensamiento y metáfora a los sistemas de información da como resultado la creación de interesantes metáforas complejas.

Puede señalarse tres componentes principales en los sistemas de información (SEGIET y BROUSE, 1992:10-11), que corresponden a los datos, el tratamiento, las reglas y normas, y los procesadores:

1. Información.
2. Usuarios.
3. Elementos soporte: vehículos de información, procesadores de información, difusores de información y almacenes de información.

A su vez, las características de los sistemas de información pueden agruparse en:

1. Tecnológicas: rendimiento y seguridad desde el punto de vista del equipo físico.
2. Funcionales y semánticas: si el sistema hace lo que debe de forma correcta y es capaz de adaptarse a requisitos cambiantes.
3. Económicas: coste del sistema y eficiencia en cumplir sus objetivos.
4. Sociales: impacto sobre el entorno social.

El objetivo del sistema de información, y de sus componentes, es la gestión de información de todo tipo. Por ello, busca facilitar el desempeño de las actividades administrativas, de gestión y de producción en todos los niveles de la

organización, lo que realiza suministrando la información adecuada, con la calidad adecuada, a la persona adecuada, en el momento y lugar oportunos, y con el formato preciso para que su uso sea correcto. Para alcanzar los objetivos previstos, el sistema debe ser capaz de desempeñar las siguientes funciones:

1. Recogida de datos de entrada.
2. Evaluación de calidad y relevancia de los datos.
3. Manipulación o proceso de los datos.
4. Almacenamiento de la información.
5. Distribución de la información.

Normalmente, los datos que manipule un sistema de información pueden dividirse en tres grandes categorías, que corresponden a microdatos (datos rutinarios, desagregados, sin influencia en la decisión), y metadatos (información y datos sobre los propios datos y su significado).

9.4.2. Tipos de sistemas de información.

Independientemente de los componentes y de las funciones, se han realizado numerosos esfuerzos por ofrecer una tipología de los sistemas de información, algunas de las cuales se citan seguidamente. Por una parte, se han señalado tres tipos de posibles sistemas de información dentro de una organización (SENN: 25-30):

1. Sistemas de procesamiento de datos: se desarrollan para procesar grandes volúmenes de datos generados en actividades rutinarias, que normalmente siguen procedimientos de operación estándar. También se les llama sistemas para procesamiento de transacciones.
2. Sistemas de información para la administración, o de información administrativa (*MIS, Management Information System*): son aquellos que, accediendo a los anteriores, ofrecen información que se utiliza para la toma de decisiones y el control en situaciones bien estructuradas, en las que puede utilizarse otro tipo de informaciones.
3. Sistemas de apoyo a las decisiones (*DSS, Decision Support Systems*) : ofrecen información personalizada para el usuario determinado, cuando debe tomar una decisión ante situaciones particulares semiestructuradas o no estructuradas.

De acuerdo a las decisiones, como ya se ha señalado en el párrafo anterior, se pueden dividir los sistemas en los que son sistemas de decisión estructurada y sistemas de apoyo para decisiones semiestructuradas, de acuerdo con el tipo de apoyo que prestan a la actividad de la organización:

- Sistemas de decisión estructurada: son aquellos que pueden automatizarse si se identifican las condiciones. Suelen utilizar árboles y tablas de decisión, que cubren todas las posibles situaciones.
- Sistemas de apoyo para decisiones semiestructuradas: suelen identificarse con los DDS, ya señalados anteriormente. Su principal

función es organizar la información que se utilizará para tomar una decisión, e incluyen la consideración de la existencia de un factor de riesgo y de criterios múltiples.

9.4.3. Hacia un ciclo de vida de los sistemas de información.

Dentro de las diferentes metodologías de desarrollo desempeña una función fundamental la creación de modelos. Estos modelos permiten la comprensión del funcionamiento del sistema, descubrir problemas y solucionar errores, durante el proceso de desarrollo, con el ahorro de recursos que ello conlleva. La utilización de los modelos facilitan, de la misma forma, la automatización de las funciones del sistema considerado. Los prototipos son creados utilizando como base los modelos lógicos establecidos, y, si resultan válidos, se incorporan al sistema final.

Un sistema de información, y todo el complejo proceso de su creación y desarrollo, debe obedecer a unas justificadas razones de la organización. Senn (SENN, 1992) ha propuesto un resumen de las cinco principales razones que motivan el desarrollo de nuevos sistemas:

1. Capacidad: mayor velocidad de procesamiento e incremento en el volumen de datos, así como una recuperación de la información más rápida.
2. Control: mejora y exactitud en la consistencia de la información manejada por la organización.
3. Comunicación: mejora de la comunicación entre las diferentes unidades de la organización, así como una mayor integración.
4. Costes: monitoreo y reducción de costes.
5. Ventaja competitiva: una amplia consideración en la que se incluye atraer clientes, superar a la competencia, lograr mejores acuerdos con los proveedores y desarrollar nuevos productos.

El ciclo de vida del sistema debe verse desde la perspectiva de la mejora para optimizar el sistema, de acuerdo a las necesidades de cada momento. Esto se consigue gracias al pilotaje o control del sistema, que reúne las funciones de seguimiento, análisis y reacción, si es necesario, y se asegura de que el sistema ejecuta los planes previstos..

9.4.4 Hacia una planificación de los sistemas de información.

Por último, no hay que olvidar que un sistema de información ha de responder a unas políticas bien definidas de información para la organización, acordes con su finalidad. Estas políticas desempeñan una función de marco y relación entre las partes internas y externas de la organización, y, por ende, su importancia es elevada.

La importancia que tiene el sistema de información para la organización o el proyecto que lo ejecuta obliga a pensaren una planificación a medio y largo

plazo de los sistemas de información, en lo que se ha llamado “planeación estratégica de los sistemas de información”. En realidad, esta planeación es la primera fase del diseño de un sistema de información, y su finalidad es establecer el marco de la política de información en la organización, así como hacer efectivo el entendimiento entre los diseñadores y los usuarios de los sistemas. Las etapas de la planeación estratégica son (BURCH y GRUDNITSKY, : 591-599):

1. Establecer las metas de los sistemas de información.
2. Determinar y asignar prioridades a las solicitudes de proyectos de sistemas de información
3. Determinación de los recursos y la capacidad de los sistemas de información.

9.4.5. La evolución de los sistemas de información

Una vez resuelta la fase de planeación estratégica, se utiliza la metodología de desarrollo de sistemas para crear el sistema de información pertinente. La disciplina que se encarga de los sistemas se ha denominado ingeniería de sistemas (IS), y se concibe como la encargada de planificar, diseñar y operar los sistemas, utilizando técnicas de investigación operativa. La propia dinámica sistémica en el desarrollo de sistemas de información propicia la innovación y la aparición de nuevos enfoques. Por ejemplo, la evolución natural de los sistemas de información ha llevado a poder clasificar éstos dentro de uno de los cuatro tipos siguientes, superando otras clasificaciones anteriores:

- TTS, o *Traditional Transaction Systems*: tratan información repetitiva, como nóminas, contabilidad, pedidos, etc, buscando velocidad, eficacia o control de datos masivo. Se apoyan en bases de datos y en desarrollos estructurados y/o modulares.
- DDS, o *Decission Support Systems*: analizan la información de los anteriores, dando como resultado información elaborada para facilitar la toma de decisiones. Se basan en herramientas estadísticas, utilizando lenguajes de cuarta generación e interfaces gráficos de usuario. Uno de sus derivados son los EIS (Executive Information Systems), orientados a la alta dirección.
- OIS, u *Office Information Systems*: se encargan de distribuir la información que es generada por los TTS y tratada por los DSS. Utilizan bases de datos distribuidas, multimedia, herramientas ofimáticas…
- BPA, o *Business Process Automatical*: cubren aspectos más generales de las funciones de la organización, sirviendo como apoyo a las tareas de organización, planificación y decisión, así como para el control. Una característica añadida muy importante debe ser su capacidad para adaptarse rápidamente a las cambiantes situaciones competitivas.

Superados los estadios expuestos, es el concepto de integración de sistemas de información, que comprende la interrelación de la arquitectura de información, de los sistemas abiertos y de la interconexión, la que caracterizará la nueva generación de sistemas de información. Integración que, con la misma

filosofía, irá borrando las limitaciones en las combinaciones de información de gestión, factual y documental.

9.5. Sistemas de información documental: definición, características y elementos.

Es necesario terminar concluyendo con una primera definición de los sistemas de información documental. Se pueden definir como un tipo especial de sistema de información, cuyo objeto de trabajo es el documento, de cualquier tipo, sobre el que se realiza un tratamiento documental conforme a las técnicas establecidas por los principios generales de las Ciencias de la Documentación, para resolver los problemas planteados en una organización dada. Dos serán las funciones principales que deberá resolver el sistema de información documental: la representación de los documentos, lo que hará a través de la clasificación y técnicas complementarias, y la recuperación de los documentos.

Los elementos técnicos que formarán el sistema de información documental serán los comunes a los sistemas de información en su sentido más amplio. Si se debe insistir en que, desde nuestra perspectiva, el tipo de sistema de información al que más se asemeja el sistema de información documental es el denominado DSS (*Decission Support System*, o Sistema de Soporte a las Decisiones), en cuanto se trata de un sistema que debe hacer frente a situaciones poco o nada estructuradas, y debe ser capaz de ofrecer respuestas concretas a necesidades vagas o imprecisas, que irán modificándose durante el propio desempeño del sistema. Esto no debe ser óbice para considerar que también puede darse un tratamiento documental en situaciones perfectamente estructuradas, como es el caso de la recuperación de documentos a través de una clasificación monojerárquica. No se puede admitir la pretendida probabilidad del sistema, ya que esa probabilidad es resultado de la introducción o recuperación de datos extrínsecos, no de los mecanismos intrínsecos del sistema. Los sistemas de información documental deben superar el concepto tradicional del texto como objeto de trabajo, e incorporar nuevos tipos de información documental, como la gráfica o la numérica, así como integrar las pautas y normas de trabajo, asegurando, por otra parte, el acceso a repositorios de información heterogéneos en espacio y tiempo.

Página en blanco

10. INFORMATIZACIÓN DE UNIDADES DE INFORMACIÓN. AUTOMATIZACIÓN DE BIBLIOTECAS.

Lecturas recomendadas: ORERA, 1996; MANDELBAUM, 1992; HOLT, 1987; CLAYTON, 1991; CLAYTON y BATT, 1992; MOYA, 1994; GREDLEY Y HOPKINSON, 1990; HILDRETH, 1898; GARCÍA MARCO, 1995b; DUVAL y MAIN, 1992; WRIGHT, 1995; SOCADI, 1992; BIBLIOTECA NACIONAL DE ESPAÑA, 1996; JACQUESSON, 1995;

10.1. Las unidades de información como sistemas de información.

Las unidades de información, bien sean bibliotecas, archivos o centros de información y/o documentación, responden, en su más amplia acepción, al concepto de sistema de información. El concepto de sistema de información es polisémico, es decir, posee múltiples significados e interpretaciones, dependiendo del contexto en el cual es utilizado. Para el área de interés que ocupa este capítulo, se va a considerar la biblioteca como un sistema de información.

Esto es posible ya que la biblioteca trabaja, en todo momento, no con libros, sino con información (ORERA, 1996). La biblioteca, como intermediario entre el usuario y la información, realiza las funciones de identificar, clasificar y facilitar el acceso y recuperación a la información contenida en soporte libro, sin perjuicio de que pueda además hacerlo con información en otros soportes. Todos los procesos que se realizan en una biblioteca tienen como principal elemento la información: los procesos de catalogación permiten obtener información para facilitar el acceso al original, el proceso de préstamo genera información sobre uso y usuarios, etc. Por lo tanto, se puede identificar una biblioteca con un sistema de información, en cuanto se trata de un conjunto de elementos altamente interrelacionados, y cuyas relaciones se basan casi en su totalidad en el intercambio de información entre ellas (CORBIN 1985). Como sistemas de información, las entradas al sistema son siempre de información o de peticiones de información, y sus salidas son siempre información.

Como se ha visto con anterioridad, los sistemas informáticos son también un tipo de sistemas de información. Por lo tanto, puede ser posible encontrar alguna similitud o semejanza entre ellos, lo que facilita la integración de ambos tipos. Esta premisa es la que fundamenta la utilización de sistemas informáticos en el proceso informativo que se realiza en la biblioteca. Y esta premisa también permite considerar la informatización de otros tipos de unidades de información, como archivos y centros de documentación.

10.2. La informatización de unidades de información.

Las numerosas y complejas tareas que desarrollan las unidades de información, cualquiera que sea su tipo, encuentran en la información (sea documental, procesual o de gestión) su materia de trabajo (LUCAS, 1987). Como principio general, se debe aceptar que en las unidades de información pueden identificarse dos flujos informativos principales: por una parte, la información documental, objeto básico de trabajo de la unidad, y por otra la información procesual o de gestión (los términos utilizados varían según autores), correspondiente a los procesos y a la gestión de la unidad, generada como consecuencia de las tareas de control y acceso a los documentos, la gestión económica y de usuarios, etc (CLAYTON, 1991).

Identificados ambos flujos de información, cada uno de ellos con sus particularidades propias, pueden encontrarse entre ambos un nexo básico, imprescindible: se trata de la acción que relaciona ambos flujos, y que corresponde a una de las funciones clásicas de la unidad de información: el acceso y difusión de la información documental. Por ejemplo, en una biblioteca puede distinguirse fácilmente el proceso técnico de tratamiento de los fondos, del control de préstamo y de los usuarios. El proceso técnico tiene flujos propios de información, y existe independientemente de los usuarios. Éstos, a su vez, utilizan determinada información para solicitar un ejemplar y disponer de él un tiempo dado, en un proceso independiente del tratamiento técnico. Ambos flujos pueden relacionarse, y generalmente así sucede, cuando el usuario consulta el catálogo.

Los dos flujos de información, y por extensión los sistemas que conforman, son suceptibles de informatización, y como consecuencia, también la relación existente entre ellos. Aceptado este principio, dependerá de la política de la unidad de información el adoptar o no un sistema informático, el nivel de cobertura del mismo, las características deseables... Como puede deducirse, el especialista en información puede encontrar sistemas totales, parciales, utilizar sistemas preconstruidos, o puede optar por desarrollar los suyos propios, adaptar sistemas de terceros o combinarlos según su criterio (DUCHEMIN, 1996). Las siguientes páginas toman como ejemplo y analizan las características ideales de un sistema de informatización de bibliotecas, por tratarse de una de las áreas en las que más se ha desarrollado la informatización de unidades de información (JACQUESSON, 1995). Esto no debe ser óbice para considerar que los principios expuestos en los párrafos anteriores son de aplicación a cualquier unidad de información, considerando siempre el entorno o contexto de la misma.

10.3. Los sistemas de informatización de bibliotecas.

Los sistemas de informatización de bibliotecas ya han sido definidos con anterioridad como un tipo particular de los sistemas de gestión documental, especializados en un ámbito concreto de actividad. Es preciso puntualizar que, en primer lugar, los sistemas de informatización de bibliotecas pueden ser parciales o integrados.

Por parciales deben entenderse aquellos sistemas que sólo prestan soporte informático a parte de las actividades informativas desarrolladas en la biblioteca. Un sistema que gestionase exclusivamente el proceso de préstamo, o que sólo se utilizase para catalogación, podrían servir como ejemplos de estos sistemas. Sistemas integrados son aquellos que ofrecen soporte a todos los procesos informativos desarrollados en la biblioteca, integrando y compartiendo la información entre ellos. Por ejemplo, un sistema integrado poseería subsistemas de catalogación, préstamo, adquisiciones, etc, capaces de intercambiar información entre ellos. Un análisis detallado de los mismos, pero bajo el concepto de sistemas híbridos, ha sido realizado por MOYA (MOYA ANEGÓN, 1994), que describe el proceso desde la planificación hasta los mecanismos de recuperación, pasando por la definición y normalización de las estructuras de datos necesarias. Un sistema de este tipo adopta un SGBD relacional para el control y manipulación de datos de gestión (adquisiciones, préstamo, control de usuarios...), y un SGBDD o un IRS para la gestión del catálogo, siendo ambos sistemas capaces de intercambiar información entre sí.

Por último, hay que considerar la posibilidad de utilizar de forma parcial un sistema integrado. Es decir, poner en marcha únicamente determinados subsistemas, o bien, en aquellos sistemas que puede adquirirse por subsistemas, implantar sólo alguno de los posibles.

10.4. Funciones técnicas de un sistema de informatización de bibliotecas.

La explicación de las funciones técnicas que debe desempeñar un sistema de informatización de bibliotecas se va a realizar bajo la consideración de que se trata de un sistema integrado, es decir, que responde a todos los procesos informativos de la biblioteca de modo completo e integrado. Las funciones técnicas a las que debe dar soporte el sistema son (WRIGHT, 1995; DUVAL y MAIN, 1992):

A) *Catalogación y entrada de datos.*

1. Selección, creación y corrección de los registros bibliográficos para todo tipo de materiales. Acceso directo a toda la base de datos, permitiendo la búsqueda automática con las siguientes posibilidades:
   - buscar dentro de un campo por cualquier cadena de caracteres específica.
   - incluir los números de control, autor y título como claves de búsqueda fundamentales.

- determinar los tipos de material sobre los que efectuar la búsqueda.
- utilización de operadores booleanos y de otros tipos.

2. Selección y transferencia de registros bibliográficos procedentes de fuentes externas a los ficheros locales, con funciones para conversión automática al formato propio, localización de registros duplicados, hallazgo de la presencia de etiquetas previamente definidas, verificación de la validez de las etiquetas y comprobación de la adecuación de los encabezamientos.
3. Creación de registros originales para el catálogo.
   - localización de registros duplicados.
   - acceso a ficheros de autoridades.
   - formatos de pantalla deben facilitar la entrada de las etiquetas que se precisan para los dos niveles bibliográficos (general y local).
   - validación de los números de control de ISBN e ISSN.
4. Corrección de registros originales.
   - introducción de campos y subcampos.
   - adición, rectificación y supresión de caracteres o cadenas de caracteres.
   - eliminación de registros, campos y subcampos.
5. Mantenimiento de ficheros de autoridades:
   - definición de ficheros para distintos tipos de encabezamiento (autores, entidades, materias...)
   - validación de registros introducidos.
   - creación de ficheros de autoridades con las siguientes características: estructura de referencia interrelacionada (encabezamientos autorizados, encabezamientos no autorizados y encabezamientos autorizados relacionados); supresión automática de las referencias; posibilidad del cambio total de las entradas.

B) *Control de adquisiciones y publicaciones periódicas.*

1. Capacidad de gestionar todo tipo de materiales bibliográficos.
2. Posibilidad de integrar los sistemas de control de publicaciones periódicas y de adquisiciones de monografías.
3. Reflejar los posibles métodos de adquisición: pedidos en firme, pedidos a examen, pedidos permanentes, suscripciones, donativos, canjes.
4. Acceso a los registros de pedidos y a los relacionados a través de consultas sobre pedidos concretos, proveedores y solicitantes, e informes sobre los pagos y las deudas.
5. Utilización de claves de acceso por número de pedido, ISBN e ISSN, autor, título, nombres y/o códigos de los proveedores y solicitantes.
6. Preparación del pedido: recepción de solicitudes, verificación del estado de la solicitud.
7. Creación y envío del pedido, permitiendo la entrada de los datos concretos del pedido, controlar el ajuste de los fondos y el cambio de moneda, y facilitar el envío por métodos tradicionales o mediante conexiones con los sistemas de los proveedores.
8. Gestión del pedido y suscripciones: crear, revisar, renovar, rectificar y cancelar.
9. Recepción de los documentos o partes: donativos, números especiales o suplementos. Autorización para la remisión de fascículos. Registro del suministro total o parcial de los pedidos en varios volúmenes o de los pedidos de ejemplares

múltiples. Verificación de la recepción. Control de los ajustes a los fondos presupuestarios. Devolución de volúmenes dañados o incorrectos.

10. Reclamaciones: control de pedidos incumplidos y documentos no recibidos; generación y examen automático de las reclamaciones, considerando las frecuencias de las publicaciones periódicas.

11. Correspondencia con el proveedor: envío y recepción de correspondencia; comunicación por medios impresos y electrónicos.

12. Procesamiento de facturas y pagos.

13. Mantenimiento de ficheros de desideratas.

C) *Encuadernación y microfilmación.*

Control de los documentos susceptibles de encuadernación o filmación, atendiendo a los cuatro proceso siguientes: 1. Selección; 2. Preparación; 3. Envío; 4. Devolución.

D) *Circulación.*

1. Gestión de las operaciones del préstamo.
    - todo tipo de materiales bibliográficos.
    - control automático de los documentos y de los lectores.
    - métodos de captura automática de datos (código de barras).
2. Préstamo.
    - asignación automática de plazos (rotativa, por períodos fijos).
    - control de las normas de préstamo.
    - capacidad de préstamo anterior a la introducción del registro bibliográfico en el sistema.
3. Cancelación.
    - control sobre los documentos y el rango del lector.
    - cálculo de sanciones.
4. Renovación.
    - mantener un procedimiento similar a los utilizados para el préstamo, incluyendo la posibilidad de realizar renovaciones a distancia.
5. Gestión económica: sanciones; facturas relativas a la sustitución de documentos perdidos o dañados por el lector.
6. Reserva de documentos para el préstamo.
    - establecer reservas conforme a diversos criterios.
    - cancelación de reservas.
    - control de fechas de vencimiento y vigencia.
    - mantenimiento del orden de petición para las obras más solicitadas.
    - mantenimiento de registros para la reserva.
    - envío de avisos.
7. Consultas.
    - acceso en línea para la consulta de los ficheros bibliográficos, indicando: fondo existente, la disponibilidad del fondo,(en préstamo, encuadernación, perdido...).
    - situación de la demanda (reservas pendientes).
    - acceso en línea para la consulta de los lectores (información sobre el registro total o parcial del lector, detalles de la tramitación, información del rango de la biblioteca).

- protección de la información sobre el lector contra empleo no autorizado.

8. Actualización de los registros de los lectores (añadir, rectificar, borrar) mediante método directo o diferido.
9. Control por parte de las sucursales del sistema del intercambio o transferencia del préstamo.

E) *Préstamo.*

1. Otras cuestiones a considerar respecto a la transmisión de información, como la integración de otros métodos de comunicaciones (modem, fax...), utilización de formularios normalizados para el préstamo interbibliotecario, o transmisión directa de la información solicitada a través de redes de telecomunicaciones.
2. Los ficheros de préstamo deben suministrar una información completa sobre el préstamo, determinar el estado actual de un documento concreto en el ciclo de solicitud y proporcionar detalles de las solicitudes/tramitaciones pendientes de un determinado usuario.
3. Control del proceso de entrada de solicitudes por los usuarios de la biblioteca, a través de:
    - acceso completo a los datos ya creados en otras partes del sistema (catálogo, adquisiciones, lectores).
    - diversas opciones para elegir el destino de una solicitud.
    - medios para imprimir los formularios de préstamo interno.
    - utilización de métodos de transmisión electrónica.
4. Control del proceso de entrada de solicitudes por los usuarios de la biblioteca.
    - acceso completo a los datos existentes en el sistema.
    - creación de registros para el préstamo presentes en otras áreas del sistema y se consideren apropiados.
5. Control del proceso de entrada de documentos para satisfacer las solicitudes de un usuario, ofreciendo los medios para registrar los datos necesarios para controlar la circulación del documento, y otras condiciones que se puedan agregar al documento o al lector.
6. Control del proceso de documentos en préstamo o devueltos por bibliotecas externas o usuarios internos (cancelación, renovación, elaboración de avisos y de reclamaciones).
7. Gestión de solicitudes no satisfechas, revisando su resolución mediante fuentes externas en un plazo determinado.

F) *Control del presupuesto y de la contabilidad.*

1. Proporcionar y mantener con seguridad la información que permita el control y el examen del presupuesto total.
2. Posibilidad de recurrir a la información seleccionada en otras áreas (p.ej. adquisiciones) para utilizarla conjuntamente con otra, pero sin sustituir los controles independientes y más específicos que se puedan realizar sobre esas áreas.
3. Además de proporcionar información para uso local, debería tener capacidad para manipular los formularios que se requieran para informar a los órganos ejecutivos.
4. Permitir una descripción que refleje la estructura de los capítulos de un presupuesto.

5. Permitir la introducción y corrección de los datos pertinentes para la transferencia desde otros sistemas así como las entradas directas.
6. Proporcionar resúmenes de información en forma de salida impresa, según se solicite, o a intervalos establecidos.

G) *Salidas de información*

1. Salidas habituales del sistema: apoyo a los módulos de circulación y de adquisiciones (peticiones de libros, avisos de vencimiento, subsistema de catalogación...)
2. Funciones ofrecidas por el catálogo de acceso público en línea:
   - funciones de ayuda.
   - consulta de la base de datos en general o a las partes que se seleccionen.
   - capacidad para buscar todos los términos del índice o los que se especifiquen.
   - disponibilidad de información de fondos potenciales o actuales, presente en otros módulos del sistema.
   - capacidad para salvar los resultados de búsquedas para posteriores tratamientos.
   - examen por los lectores de sus propios registros y detalles de las tramitaciones, y opciones para reservas directas.
3. Técnicas disponibles en el catálogo de acceso público en línea:
   - limitación de las búsquedas de términos mediante varios factores calificativos.
   - truncamiento de términos.
   - utilización de operadores booleanos.
   - lectura completa de los índices y los conjuntos de datos recuperados.

H) *Acceso a sistemas externos de recuperación de la información.*

1. Capacidad para conectar los sistemas de apoyo técnico para la biblioteca con los sistemas externos de recuperación mediante red y uso de telecomunicaciones. Debería ser posible acceder a los sistemas externos conectado a través de un procedimiento común. La selección de la opción de uso de un sistema externo debe dar como resultado el acceso al mismo.
2. Capacidad de limitar la búsqueda conforme a parámetros especificados y la utilización de operadores booleanos
3. Capacidad de cargar los registros seleccionados en los sistemas externos, con objeto de permitir almacenamiento y comparación con los registros locales.

10.5. El Catálogo Público de Acceso en Línea.

El OPAC (*Online Public Access Catalog*, o Catálogo Público de Acceso en Línea) es la interfaz que mantiene un sistema de automatización de bibliotecas con el usuario final. A través de la presentación de información y de las funciones ofrecidas por el OPAC, el usuario puede consultar los fondos disponibles en la biblioteca (o en otras bibliotecas cuyos catálogos se encuentren accesibles) conocer el estado de la copia, realizar reservas, revisar el estado de los préstamos

que le competen, etc. Se trata, en definitiva, del elemento de interacción y participación del usuario en el proceso informativo que se desarrolla en la biblioteca, en el que desempeña un papel fundamental el proceso comunicativo. Relacionadas con la propia evolución de los sistemas informáticos se ha establecido tres fases clásicas de desarrollo en los OPACs (HILDRETH, 1989), que pueden delimitarse como fase de línea de órdenes, fase de menú de elementos y fase de entorno gráfico de usuario.

El OPACs ofrece al usuario un subconjunto de las funciones de tratamiento de la información que se ejecutan en el centro. Ofrece (o debe ofrecer) una impresión de integración, aislando al usuario de la máquina y de las estructuras de datos que subyacen al sistema de automatización de bibliotecas. El interfaz es el medio de comunicación con el catálogo (DUVAL y MAIN, 1992), que se interroga utilizando el lenguaje propio del sistema, o utilizando sistemas de ayuda e interrogación asistida, gracias a los cuales el usuario se ve libre de las complicaciones de todo lenguaje de interrogación. Deben destacarse las funciones de acceso y consulta directa por materias o autores como uno de los elementos básicos a utilizar en un OPAC. Todos los sistemas de informatización de bibliotecas ofrecen un módulo de OPAC. Sin embargo, el diseño y las funciones presentes en cada uno de ellos pueden variar, a causa de los intereses y decisiones de los diferentes fabricantes. De hecho, la mayor parte de los interfaces son diferentes, aunque den soporte a funciones similares, lo que supone un inconveniente para el usuario. Ante esta situación, cobran especial importancia los factores y elementos de amigabilidad que incorpore la interfaz para el usuario (GARCÍA MARCO, 1995b).

Como corolario, cabe afirmar que un OPAC ofrecerá un pobre rendimiento si la biblioteca no considera la formación de los usuarios como un factor clave de éxito. La asimilación, por parte de los mismos, del uso del OPAC como una herramienta diaria, de sus capacidades y de sus límites, es un elemento inherente al propio proceso de informatización de centro, que de ningún modo puede ser obviado o pospuesto frente a otros intereses.

10.6. El proceso de informatización.

La selección e implantación de un sistema de informatización de bibliotecas resulta ser una cuestión vital para el funcionamiento bibliotecario. Por lo tanto, todo el proceso a seguir debe intentar eliminar y minimizar cualquier posible y/o futuro problema, normalmente resultado de un incorrecto proyecto de informatización. Se hace imprescindible una rigurosa planificación y estudio del entorno y de la implantación del sistema, acorde con el significado que debe tener, para una biblioteca, la introducción y puesta en marcha de un sistema de este tipo (MANDELBAUM, 1992).

No todas las situaciones en las que puede encontrarse una biblioteca resultan ser idóneas para la implantación de un sistema informático. resulta necesario un riguroso y documentado estudio previo sobre la situación de la biblioteca, su contexto y sus procedimientos de trabajo. Es posible afirmar que existe una relación directa entre la calidad de la planificación y el éxito de la

implantación y puesta en marcha del sistema informático. En este particular, debe ocupar un lugar principal el estudio económico que analice las consecuencias económicas de la implantación de un sistema informático. No puede admitirse, como se afirma en ocasiones, que no es posible cuantificar económicamente la rentabilidad de la instalación. Si el estudio ha sido realizado correctamente, pueden utilizarse numerosos indicadores económicos (coste actividades y horas de trabajo, etc.), que demuestren la rentabilidad de la opción. Esto exige honestidad en la aceptación de una conclusión en contra, que puede verse modificada por razones de política bibliotecaria o de otro tipo.

Al aceptar anteriormente la configuración como sistema de la biblioteca, así como de la aplicación informática, es posible adoptar, a su vez, un enfoque sistémico para el proceso de informatización de la biblioteca, cuyas fases han sido señaladas por Clayton (CLAYTON, 1991; CLAYTON y BATT, 1992) :

1.- Plan de desarrollo global: el plan de desarrollo global debe establecer las fases del proceso, la duración de cada una de ellas, los resultados esperados y el personal responsable. Este documento debe fijar la estrategia general del proyecto de informatización.
2.- Análisis de sistemas: esta fase debe desmenuzar, hasta sus niveles más detallados, el funcionamiento del sistema bibliotecario existente, que va a ser objeto de informatización. Estudiará todos los procesos, recursos y normas existentes en el sistema, sus interacciones... Detallará los aciertos y errores del mismo, así como las acciones a poner en marcha para lograr su optimización. El resultado de esta fase será un completo y exhaustivo documento en el cual se especificaran todas estas cuestiones.
3.- Especificación de los requisitos o requerimientos: esta es una parte crítica del proyecto. Deben detallarse todas y cada una de las características y/o prestaciones que debe ofrecer el sistema informático que se considera ideal para su implantación en la biblioteca. La base de todos estos requerimientos serán las necesidades que la biblioteca debe satisfacer para alcanzar el conjunto de objetivos que tiene previstos.
4.- Evaluación de sistemas: con la óptica del documento que contiene los requerimientos, deben contemplarse las opciones que se ofrecen para la selección de un sistema informatizado. La biblioteca puede optar por desarrollar uno propio (lo cual resulta casi imposible en la mayor parte de las situaciones), o adquirir alguno de los existentes en el mercado, fabricados por terceras partes. En este segundo caso, se remite a los proveedores el documento de requerimientos, a la espera de sus propuestas. Las respuestas de éstos, así como el análisis de la literatura especializada, y la visita e intercambio de experiencias con otros usuarios, serán la base sobre la que tomar una decisión.
5.- Negociación del contrato y de sus condiciones: la selección de un sistema debe seguirse por el contacto con el proveedor para negociar las condiciones del contrato de compra. En este caso puede darse una variada casuística, por lo que es recomendable atender, cuando menos, a condiciones referidas a fechas de puesta en marcha, formación de usuarios, mantenimiento y pruebas de aceptación, estado de propiedad de las patentes, formas de pago, etc.

6.- Puesta en marcha: la fase de puesta en marcha resulta ser la culminación de proyecto. Puede ser parcial (en sucesivas fases), o completa, lo que dependerá de la situación bibliotecaria. Sería interesante disponer de un tiempo de prueba real, antes de la aceptación total. En este contexto debe situarse la formación del personal de la biblioteca en relación con el sistema, a diferentes niveles. Deben considerarse, además, la legislación que puede afectar al ámbito de trabajo y a los operarios del sistema.
7.- Control y mantenimiento: el sistema necesitará mantenimiento y control de las actividades que desarrolla, para su perfecto funcionamiento y desempeño de sus funciones. Es necesario considerar que el mantenimiento cubierto en los contratos de adquisición suele cubrir el primer o primeros años, corriendo el gasto a cuenta de la biblioteca posteriormente.
8.- Renovación: los sistemas de informatización de bibliotecas no son eternos ni perfectos, por lo que resulta ineludible considerar que será necesaria su renovación y/o sustitución pasado determinado período de tiempo.

Resulta forzoso hacer dos últimas consideraciones. En primer lugar, la implantación de un sistema de este tipo tiene costes ocultos (como resulta en ocasiones la necesidad de modificar las condiciones ambientales de una sala), derivados normalmente de una estudio o planificación incompleto. Es necesario reducir estos costes a su mínima expresión. Por contra, son numerosos los beneficios intangibles que una informatización puede ofrecer a la biblioteca, y que deben estudiarse y explicarse detenidamente, como complemento a la justificación necesaria para la instalación de un sistema informatizado.

Para terminar este apartado debe hacerse una última recomendación. Es altamente recomendable que la biblioteca genere documentación de usuario propia sobre las funciones, procesos y normas del sistema informático. La documentación así generada, aunque puede resultar un trabajo añadido, es de gran utilidad una vez que el sistema ha sido puesto en marcha. Es evidente que la disponibilidad de manuales asequibles para el personal y los usuarios del sistema harán más fácil la comprensión e integración del mismo en el entorno bibliotecario.

10.7. Normas de tratamiento, recuperación, transferencia y almacenamiento de información.

Las múltiples entidades y categorías de datos que son objeto de tratamiento, en el marco de las tareas propias de una unidad de información, se ven reflejadas en los sistemas informáticos que se utilizan en la misma. Como puede imaginarse, resulta necesaria la utilización de formatos que respeten determinadas reglas, en aras de poder compartir los datos entre diferentes unidades (GREDLEY y HOPKINSON, 1990). Las agencias nacionales e internacionales de normalización han publicado normas generales que rigen las características, estructura y componentes de muchos de estos elementos (AENOR, 1994). La CDU, el préstamo interbibliotecario (ILL, *InterLibrary Loan*), o el intercambio de información entre bibliotecas y proveedores, a través de

proyectos EDI (*Electronic Data Interchange*), son actividades que han sido escrupulosamente reglamentadas mediante este tipo de normas. ILL ha sido definido a través de las normas ISO 10160 y 10161, EDI y EDIFACT mediante las normas ISO 9735 y 7372... De especial importancia es la colaboración, cada vez más extendida, entre diferentes bibliotecas que forman, agrupadas por criterios o áreas de interés, redes de bibliotecas, interconectadas usando las actuales redes de telecomunicaciones, con vistas a compartir las laboriosas tareas de catalogación, mecanismos de préstamo interbibliotecario... Mención especial merece el protocolo ANSI/NISO Z39.50, que establece las normas para facilitar una única interfaz de comunicación e interrogación de bases de datos, en relación con las normas ISO 10162 y 10163 sobre SR (*Search and Retrieval*, Búsqueda y Recuperación). Estas normas son las que deben estableceer, en un futuro próximo, los requisitos y mecanismos de interrogación de bases de datos mediante redes de comunicaciones, especialmente a través de Internet y sus sucesores, encontrándose ya en aquélla aplicaciones que hacen uso de estos protocolos, aunque no todo lo extendidos como sería deseable.

Resulta evidente que la aplicación de estas normas en el entorno de los sistemas de informatización de unidades de información redunda en todo tipo de ventajas en el momento, casi siempre necesario, de intercambiar información con otras unidades. Es sobre esta necesaria normalización sobre la que deberá asentarse el intercambio en los próximos años, parejo con el desarrollo de las redes de telecomunicaciones. Por desgracia, esta normalización no ha alcanzado todavía el nivel deseable. Por ejemplo, la aplicación de un formato común en las agencias bibliográficas y en las bibliotecas para facilitar la transferencia de registros bibliográficos, el conocido como formato MARC, tiene todavía numerosas variantes. En el ámbito archivístico el mismo esfuerzo, la definición de un formato para sus materiales, a través de un formato MARC específico, es más reciente, encontrándose todavía en sus inicios.

En los siguientes apartados se va a tratar el formato MARC para registros bibliográficos, por tratarse de uno de los más extendidos y objeto de experiencia y publicación, y corresponder, a su vez, a los sistemas de automatización de bibliotecas tratados con anterioridad.

10.8. El registro bibliográfico.

Un registro bibliográfico es un registro que ofrece información sobre una noticia bibliográfica. Este registro se compone de un conjunto de elementos que transmiten información sobre el elemento u objeto al que representan. Todos los registros de tipo bibliográfico comprenden dos tipo de elementos:

- Elementos bibliográficos: transmiten información sobre un documento .
- Elementos de procedimiento: transmiten información sobre la función de procedimiento del registro.

10.9. Breve historia del formato MARC.

El auge de los sistemas de informatización de bibliotecas, especialmente en los Estado Unidos de América, trajo consigo la posibilidad y la necesidad de utilizar un formato de registro común, de tal forma que fuese posible acometer los procesos de catalogación e intercambio de registros entre bibliotecas de una forma cooperativa. La adopción de este tipo de medidas supondría incuestionables mejoras en el sistema bibliotecario norteamericano. Como cabeza de ese sistema bibliotecario, la Biblioteca del Congreso (Library of Congress) puso en marcha un programa destinado a establecer un formato común para el almacenamiento de registros bibliográficos en el ámbito de su cobertura.

El trabajo desarrollado en la década de 1960 significó la creación y desarrollo del formato MARC (MAchine Readable Cataloging) 11, que revisado y mejorado dio lugar, en 1971, a la norma ANSI correspondiente. Esta norma fue aceptada por la ISO en 1973, que publicó una norma revisada en 1981, identificada como ISO 2709-1981. La utilización del formato MARC es inseparable de la divulgación de las normas de catalogación AACR en 1979, lo que desembocó en la redacción y adecuación de las normas de catalogación existentes en los diferentes países a la utilización de los formatos MARC. Tras la adopción de la norma en Europa, las diferentes agencias bibliográficas nacionales adoptaron el contenido de la norma a las peculiaridades de sus propios sistemas bibliotecarios, lo que ha dado como resultado gran cantidad de formatos MARC nacionales, cada uno de los cuales posee elementos y características que los diferencian del resto. Generalmente, los formatos MARC nacionales derivan del formato MARC de la Biblioteca de Congreso (USMARC) o del formato MARC de la British Library (UKMARC). Según estos criterios, resulta comprensible la aparición de formatos IBERMARC, ANNAMARC, INTERMARC, CANMARK, etc., en diferentes países. La multiplicidad reinante en el formato MARC, la cual lo alejaba de su primera misión como formato de intercambio, ha desembocado en la aparición de un nuevo formato MARC, al que se ha denominado UNIMARC, cuya finalidad es recuperar la idea de un formato MARC universal, de forma que se facilite el intercambio de registros a nivel mundial.

Inseparable de la utilización del formato MARC es la aplicación de otras normas complementarias, entre las que deben conocerse las correspondientes a la ISBD (International Standard Book Description), ISO 3297-1975 (ISSN, International Serial Standard Number), ISO 2108-1978 (ISBN, International Standard Book Number) e ISO 214-1976 (normas para la realización de resúmenes de publicaciones y documentos).

10.10. El formato MARC.

El formato MARC, tal y como es definido en la norma ISO 2709, "Es la estructura de la noticia bibliográfica, así como el sistema de códigos que identifican el contenido de la misma", "tiene como finalidad facilitar el intercambio de los datos bibliográficos en soporte magnético.", Y su función es "desarrollar el análisis de los elementos de una noticia bibliográfica, o referencia bibliográfica, especificando las etiquetas, indicadores y códigos de subcampo."

Resulta de suma importancia recordar que el formato MARC no es un formato de catalogación o una norma de análisis. La prioridad del MARC es definir un formato legible por máquina para facilitar el intercambio de registros, es decir, establece las partes y los identificadores de esas partes del registro bibliográfico, pero no establece rígidamente su disposición. Esto es lo que permite encontrar registros MARC de diferente estructura y ordenación, ya que son correctos siempre y cuando respeten la norma, y esta última sólo se refiere a la estructura física del registro (SOCADI, 1992).

El formato físico utilizado para los registros tienes tres partes:

1. La estructura del registro: es la representación física de la información en el soporte legible por el ordenador.
2. Los designadores del contenido: son las etiquetas y códigos que sirven para identificar a los mismos.
3. El contenido del registro: la propia información.

La estructura del registro posee, a su vez, tres partes, correspondientes a:

1. Cabecera: de longitud y formato fijos, informa sobre la longitud, tipo y nivel bibliográfico del documento.
2. Directorio: informa sobre los campos variables específicos incluidos en el formato.
3. Campos variables: o área de datos, contiene los datos de cada campo.

De lo anterior puede deducirse que el formato MARC posee dos tipos de campos: los de longitud fija y los de longitud variable. Los primeros se utilizan para aquellos elementos de longitud fácilmente precisable, y que siempre aparecen en el registro. Los segundos, por contra, son preferidos para campos cuyo contenido puede variar notablemente, dependiendo del nivel de la noticia bibliográfica.

10.10.1. El formato IBERMARC.

En consonancia con el panorama esbozado anteriormente, el sistema bibliotecario español utiliza el formato IBERMARC, del cual existen versiones para monografías, seriadas, registros sonoros y de imagen, materiales cartográficos, etc., de los cuales se han publicado manuales y referencias, sometidas a revisión cada cierto tiempo (BIBLIOTECA NACIONAL DE ESPAÑA, 1996). Algunas comunidades autónomas han implantado su propio formato MARC, como es el caso de Cataluña (CATMARC). Como ejemplo, el formato IBERMARC para monografías ofrece la siguientes estructura:

1. Cabecera y guía, de 24 caracteres.
2. Directorio o repertorio, número variables de elementos de longitud fija (12 caracteres).
3. Campos de longitud fija (número de identificación, campo 001, y códigos de información, campo 008).
4. Campos de longitud variable (números de control, lenguas, número de clasificación y localización, descripción bibliográfica y registro de entidades secundarias).

Un campo IBERMARC contiene un conjunto de elementos que identifican al campo, sus características, sus subcampos y los datos contenidos en él. La revisión del manual del formato IBERMARC correspondiente permite saber si los campos pueden ser repetibles o no repetibles. Por ejemplo, puede utilizarse el campo 240, correspondiente al título uniforme:

240 b0 $a[Constitución,$d1978]

240: etiqueta que identifica al campo.
b0: indicadores
$: etiqueta de subcampo
a: identificador del Título Uniforme.
Constitución: dato de Título Uniforme.
$: etiqueta de subcampo.
d: identificador del subcampo fecha.
1978: dato del subcampo fecha.

10.10.2. Intentos de integración: el formato UNIMARC.

La diversidad de formatos MARC hace difícil el intercambio de registros bibliográficos entre diferentes agencias nacionales, lo que en principio resultaba ser su principal objetivo. En realidad, el intercambio de registros supone, en la actualidad, la utilización de programas capaces de "traducir" entre diferentes formatos. Por esta razón se han iniciado esfuerzos en aras de obtener un formato MARC común, al que se ha llamado UNIMARC (HOLT, 1987). Sin embargo, la existencia de grandes bases de datos y de aplicaciones específicas construidas para los diferentes formatos hace inviable, por el momento, la implantación del formato UNIMARC. Por esta causa, se tiende a ver el formato UNIMARC, en estos momentos, más como un formato de intercambio entre diferentes MARC, que como un formato común. Es decir, que en vez de intercambiar registros entre diferentes MARC, con el esfuerzo particular que ello supone, se tiende a implantar programas que "traduzcan" el MARC propio a UNIMARC, y programas que "traduzcan" de UNIMARC al propio, en el sentido de agilizar y ahorrar esfuerzos en las tareas de transferencia. La estructura de los campos de los registros UNIMARC responde a las siguientes partes:

- 001-071: Bloque de identificación.
- 100-135: Bloque de información codificada.
- 200-225: Bloque de información descriptiva.
- 300-345: Bloque de notas.
- 410-488: Bloque de entradas relacionadas.
- 500-541: Bloque de título relacionado.
- 600-686: Bloque de análisis del sujeto o de materias.
- 700-722: Bloque de responsabilidad intelectual.
- 801-802: Bloque de uso internacional.
- 9... : Bloque de uso nacional.

.

11. BIBLIOGRAFÍA.

Abreviaturas:

ACM Association for Computer Machinery

ARIST Annual Review of Information Science and Technologiy

JASIS Journal of the American Society for Information Science.

IFID International Forum on Information and Documentation.

ABOBA, B. (1993): *The Online User's Encyclopedia*. Reading: Addison-Wesley.

ALBERICO, R. y MICCO, M. (1990): *Expert Systems for Reference and Information Retrieval*. Wesport: Meckler.

ALLEYRAND, M. R. (1989): *Image Storage and Retrieval Systems*. New York: McGraw-Hill.

APERS, P.G.M., BLANKEN, H.M. y HOUTSMA, M.A.W. (1997): *Multimedia databases in perspective*. New York : Springer.

ARMSTRONG, C. J. y LARGE, J. A. (1988): *Manual of online search strategies.* Aldershot: Gower, 1988.

ASHFORD, J.H., y WILLET, P. (1988): *Text retrieval and document databases*. London: Chartwell Bratt.

AZORÍN, V., y LÓPEZ, T. (1994): “Problemas que presenta la descripción de imágenes en la elaboración de bases de datos multimedia.” *Actas de las IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94*, Gijón: Universidad de Oviedo, pp.411-416.

BASCH, R. (1993): *Secrets of the Super Searchers*. Wilton: Eigth Books.

BATES, M.J. (1979): “Information Search Tactics.” *JASIS*, 30, 4, pp. 205-214

BATES, M.J. (1981): “Search techniques.” *ARIST*, 16, pp. 139-169

BATES, M.J. (1987): “How to Use Information Search Tactics Online.” *Online*, 11, pp. 47-54.

BATINI, C., CERI, S. y NAVATHE, S.B. (1994): *Diseño conceptual de bases de datos. Un enfoque de entidades-interrelaciones*. Wilmington: Addison-Wesley/ Díaz de Santos.

BEARD, D.V. (1991): "Computer Human Interaction for Image Information Systems." *JASIS* 42, 8, pp. 600-608.

BELKIN, N., CROFT, W.B. (1987): “Retrieval Techniques.” *ARIST*, 22, pp. 109-145

BELKIN, N., ODDY, R. y BROOKS, H. (1982): “ASK for Information Retrieval: 1. Background and Theory.” *Journal of Documentation*, 38, 2, pp. 61-71. 145-164 (part.2).

BELKIN, N., ODDY, R. y BROOKS, H. (1982): “ASK for Information Retrieval: 2. Results of a Design Study.” *Journal of Documentation*, 38, 3, pp. 145-164.

BERTINO, E. y MARTINO, L. (1995): *Sistemas de bases de datos orientadas a objetos*. Wilmington: Addison-Wesley/Díaz de Santos.

BIANCHI, N. *et alii*, (1996): “Multimedia document management: an anthropocentric approach.” *Information Processing & Management*, 32, 3, pp.287-304.

BIBLIOTECA NACIONAL DE ESPAÑA, (1996): *Formato IBERMARC para registros bibliográficos.* Madrid: Biblioteca Nacional de España, Ministerio de Cultura.

BISHOP, P. (1992): *Fundamentos de Informática*. Madrid: Anaya Multimedia.

BLAIR, D.C. (1990): *Language and Representation in Information Retrieval*. Amsterdam: Elsevier.

BLAIR, D.C. (1996): *Progress and Problems in Information Retrieval*. London: Library Association Pub.

BLAIR, D.C. y GORDON, M.D. (1991): “The management and control of written information.” *Information & Management*, 29, pp.239-246.

BLAIR, D. y MARON, M.E. (1990): “Full-Text Information Retrieval: Further analysis and Clarification.” *Information Processing & Management*, 26, 3, pp. 437-447.

BOYD RAYWARD, W. (1996): “The History and Historiography of Information Sciences: Some Reflections.” *Information Processing & Management*, 32, 1, pp. 3-17.

BUCKLAND, M. (1991a): *Information and Information Systems*. New York: Greenwood Press.

BUCKLAND, M. K. (1991b) "Information Retrieval of More than Text." *JASIS*, 42, 8, pp. 586-588.

BUSON, R. (1993): "Base de datos documentales." *PC World*, Enero, pp.120-126.

BUSTELO, C. (1994): “El papel de los expertos en documentación en la implantación de tecnologías de gestión de documentos.” *Actas de las IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94*, Gijón: Universidad de Oviedo, pp.357-362.

CANALS CABIRÓ, I. (1990): “El concepto de hipertexto y el futuro de la documentación.” *Terceras Jornadas Españolas de Documentación Automatizada Documat 90.* Palma de Mallorca: Univ. de les Illes Balears, pp.49-76.

CARIDAD, M., y MOSCOSO, P. (1991): *Los sistemas de hipertexto e hipermedias. Una nueva aplicación en informática documental*. Madrid: Fundación Germán Sánchez Ruipérez.

CHAMORRO, R., y GONZALEZ, M. (1992): *IBERTEX. El videotexto español*. Madrid: Paraninfo.

CHAUMIER, J. (1993): *Técnicas de documentación y archivo*. Barcelona: oikos-tau.

CHEN, D.P. (1976): “The entity-relationship model: Towards a unified view of data.” *ACM Transactions on Database Systems*, 1, pp.9-36.

CHEN, H. y DHAR, V. (1990): “A Knowledge-Based Approach to the Design of Document-Based Retrieval Systems.*” Proceedings of Conference on Office Information Systems, ACM SIGOIS Bulletin*, 11, 2-3, pp. 281-290.

CHORAFAS, D.N. (1994): *Intelligent Multimedia Databases*. Englewood Cliffs: Prentice-Hall.

CLAYTON, M. (1991): *Gestión de automatización de bibliotecas*. Madrid: Fund. Germán Sánchez Ruipérez.

CLAYTON, M., y BATT, C. (1992): *Managing Library Automation*. Aldershot: Gower.

CODINA, L. (1990): "Bases de datos documentales para microordenadores." *Actas de las 3as Jornadas Españolas de Documentación Automatizada DOCUMAT 90*, Palma de Mallorca: Universitat de les Illes Balears, pp. 618-627.

CODINA, L. (1994): *Sistemes d’informació documental: concepció, anàlisi i disseni de sistemes de gestió documental amb microordinadors*. Barcelona: Pòrtic.

CODINA, L. (1996): “Cómo descubrir información en Internet y cómo conseguir que nos descubran a nosotros.” *Net Conexión*, 13, pp.52-67.

CODINA, L., y ABADAL, E. (1992): "Gestió documental amb microordinadors: caracteristiques, estructura i tecnologia dels sistemes de gestió documental." *ITEM, Revista de Biblioteconomia i Documentació*, 11, pp.72-100.

CONKLIN, L. (1987): ”Hypertext: an introduction and survey.” *IEEE Computer*, 20, pp. 17-41.

CONLON, J.R. y CONLON, S.J. (1996): “Optimal Use of an Information Retrieval System.” *JASIS*, 47, 6, pp. 449-457.

CONVEY, J. (1992): *Online Information Retrieval*. London: Library Association Pub.

CORBIN, J. (1985): *Managing the Library Automation Project*. Oryx Press.

COSTANZO CAPITANI, P. (1989): *Manuale di base per il trattamento dell'informazione*. Milano: Editrice Bibliografica.

COSTILLA, C. (1992): “Avances recientes y tendencias previsibles de las bases de datos.” en MAP, *Encuentro sobre bases de datos en la Administración Pública*, Madrid, pp.119-145.

COVICE, R.A. (1994): “Using Hypermedia to Create an Exploration Space for Users.” *IFID*, 19, 3-4, pp.28-31.

COX, J. (1991): *On-line and CD-ROM Database Searching*. London: Mansell.

COYNE, J.P. (1991): "The Relationship Between Multimedia and Expert Systems.*" Multimedia Review*, June, p. 13.

CROFT, W. B. (1987): “Approaches to intelligent information retrieval.*” Information Proccesing & Management*, 23, 4, pp. 249-254

CRONIN, B., VAKKARI, P. (eds) (1992): *Conceptions of Library and Information Science.* London: Taylor Graham..

DATE, C.J. (1986): *Introducción a los sistemas de bases de datos.* México: Addison-Wesley.

DE MEY, M. (1980): "The relevance of the cognitive paradigm for information science." Harbo, O.(ed.) *Theory and Application of Information Research*. London: Mansell, pp.48-61.

DEBONS., A., HORNE, E., y CRONENWETH, S. (1988): *Information Science: An Integrated View*. Boston: G.K. Hall.

DEOGUN, J.S. y RAGHAVAN, V.V. (1988): "Integration of information retrieval and database management systems." *Information processing and management*, 24, 3, 303-313.

DESCHATELETS, G. (1986): "The Three Languages Theory in Information Retrieval." *International Classification*, 13, 3, pp.126-132

DEWIRE, D.T. (1994): *Text Management*. New York: James Martin/McGraw-Hill.

DÍAZ, P., CATENAZZI, N. y AEDO, I. (1996): *De la Multimedia a la Hipermedia*. Madrid: Ra-Ma.

DUCHEMIN, P-Y. (1996): *L'art d'informatiser une bibliothéque: guide pratique*. Paris: D. Lahary.

DUVAL, B.K., y MAIN, L. (1992): *Automated Library Systems*. *A Librarian's Guide and Teaching Manual.* Wesport: Meckler.

DVORAK, J.C y ANIS, N. (1992): *Telecomunicaciones para PC*. Madrid: McGraw-Hill.

ELLIS, D.A. (1989): "Behavioral Model for Information Retrieval*." Journal of Information Science*, 15, pp. 237-247.

ELLIS, D. (1990): *New Horizons in Information retrieval*. London: The Library Association.

ELMASRI, N., y NAVATHE, S.B. (1989): *Fundamentals of Database Systems*. Redwood City: The Benjamin/Cummings Pub.Co.

ESPINOSA, Mª. B., *et al*. (1994): *Tecnologías documentales*. Madrid: Tecnidoc.

EWERS, A. (1994): "A Review of New Developments in Text Retrieval Systems." *Journal of Information Science*, 20, 6, pp. 438-443.

FIDEL, R. *et alii* (eds.) (1994*): Challenges in Indexing Elextronics Text and Images*. Medfor: Learned Information.

FONG, J. y SIU, B. (eds) (1996*): Multimedia, knowledge-based, and object-oriented databases*. New York : Springer.

FROST, R. (1989): *Bases de datos y sistemas expertos. Ingeniería del conocimiento*. Madrid: Díaz de Santos

FUNDESCO, (1986): *La consulta a bases de datos*. Madrid: IMPI.

GARCÍA GUTIÉRREZ, A. y LUCAS FERNÁNDEZ, R. (1987): *Documentación automatizada en los medios informativos*. Madrid: Paraninfo.

GARCÍA MARCO, F.J. (1995a):“Paradigmas científicos en recuperación de información.” García Marco, F.J. (Ed.) *Organización del Conocimiento en Sistemas de Información y Documentación*. Zaragoza: ISKO, 1995, pp.99-112..

GARCÍA MARCO, F.J. (1995b): “Interfaces amigables para la recuperación de información bibliográfica.” *Scire*, 1, 1, pp.127-148.

GARCÍA MARCO, F.J. (1996a): *Tratamiento y Recuperación de la Información*. Zaragoza: Kronos.

GARCÍA MARCO, F.J. (1996b): “Vannebar Bush, el hipertexto y el futuro del documento.” Tramullas, J. (Ed.) *Tendencias de investigación en Documentación*. Zaragoza: Universidad de Zaragoza, pp.185-210.

GARCÍA MARCO, F.J. y TRAMULLAS SAZ, J. (1996): *World Wide Web: Fundamentos, navegación y lenguajes de la red mundial de información*. Madrid: Ra-Ma.

GIL LEIVA, I. Y RODRÍGUEZ MUÑOZ, J.V. (1996): “Los orígenes del almacenamiento y recuperación de información.” *Boletín de la Asociación Andaluza de Bibliotecarios*, 142, pp. 9-18.

GILSTER, P. (1996): *Finding It on the Internet*. New York: John Wiley & Sons.

GREDLEY, E. y HOPKINSON, A. (ed.) (1990): *Exchanging bibliographic data: MARC and other international formats*. London: Library Association.

GREEN, W.B. (1993): *Introduction to Electronic Documents Management*. Boston: Academic Press.

GUILERA, L. (1993): *Introducción a la informática*. Barcelona: EDUNSA.

HAHN, H. (1994): *INTERNET. Manual de referencia*. Madrid: McGraw-Hill.

HANSON, T.(ed.) (1995): *Bibliographic Software and the Electronic Library*. Hertfordshire: University of Herfordshire Press.

HAWKINS, D. T. *et alii*. (1985): "Front End Software for Online Database Searching. Part One: Definitions, System Features, and Evaluation." *Online*, November, pp. 30-37.

HEAPS, H.S. (1978): *Information Retrieval. Computational and Theoretical Aspects*. New York: Academic Press.

HERAS, A.R. DE LAS (1991): *Navegar por la información*. Madrid: Fundesco.

HEWINS, E. (1990): “Information Needs and Use Studies.” *ARIST*, 25, pp. 145-172.

HILDRETH, C.R. (ed.) (1989): *The Online Catalogue: Developments and Directions*. London: The Library Association.

HJØRLAND, B. y ALBRETCHSEN, H. (1995): “Toward a New Horizon in Information Science: Domain-Analysis.” *JASIS*, 46, 6, pp. 399-425.

HOLT, B.P. (1987): *UNIMARC Manual*. London: IFLA.

INGWERSEN, P. (1988): "Towards a new research paradigm in information retrieval." Wormell, I. (ed.) *Knowledge Engineering: Expert Systems and Information Retrieval.* London: Taylor Graham, pp. 150-168.

INGWERSEN, P. (1992): *Information Retrieval Interaction*. London: Taylor Graham.

INGWERSEN, P. (1994): "The Cognitive Perspective in Information Retrieval." *IFID*, 19, 2, pp.25-32.

JACKSON, G.A. (1990): *Introducción al diseño de bases de datos relacionales*. Madrid: Anaya Multimedia.

JACQUESSON, A. (1995): *L'informatisation des bibliothéques. Histoire, stratégie et perspectives*. Paris: Cercle de la Librairie.

JONES, K. S. (1983): "Intelligent Retrieval*." Proceedings of Informatics 7: Intelligent Information Retrieval*, London: Aslib, pp. 136-142.

KEEN, E.M. (1992): "Some aspects of proximity searching in text retrieval systems." *Journal of Information Science*, 18, 89-98.

KEMP, D.A. (1988): *Computer-based Knowledge Retrieval*. London: Aslib..

KOEGEL BUFORD, J.F. (ed.) (1994): *Multimedia Systems*. ACM Press.

KORTH, H. y SILBERSCHATZ, (1994): *Fundamentos de bases de datos*. Madrid: McGraw.Hill.

KUHLTHAU, C. (1991): "Inside the Search Process: Information Seeking from the User's Perspective." *JASIS*, 42, 5, pp. 361-371.

LANCASTER, F.W. (1978): *Towards Paperless Information Systems*. New York: Academic Press.

LANCASTER, F.W. (1986): *Vocabulary Control for Information Retrieval*. Arlington: Information Resources Press. (*El control del vocabulario en la recuperación de información*. Valencia: Universitat, 1995).

LANCASTER, F.W. y WARNER, A.J. (1993): *Information Retrieval Today*. Arlington: Information Resources Press.

LANDOW, G.P. (1995): *Hipertexto. La convergencia de la teoría crítica contemporánea y la tecnología*. Madrid: Paidos.

LASSOURY, J. (1994): *La documentique: Gestion électronique de documents et gestion documentaire*. Paris: Dunoch.

LITC, (1994): *European Directory of Text Retrieval Software*. Aldershot: Gower.

LIZASOAIN, L. (1992): *Bases de datos en CD-ROM.* Madrid: Paraninfo.

LÓPEZ YEPES, A. (1993): *Documentación Multimedia*. Salamanca: Publicaciones de la Universidad Pontificia.

LÓPEZ YEPES, J. (1995): *La Documentación como disciplina. Teoría e Historia*. Pamplona: EUNSA.

LÓPEZ YEPES, J. (COORD.) (1996): *Manual de Información y Documentación*. Madrid: Pirámide.

LUCAS, H.C. (1987): *Sistemas de información. Análisis. Diseño. Puesta a punto*. Madrid: Paraninfo.

LYNCH, C.A. (1994): "The Integrity of Digital Information: Mechanics and Definitional Issues." *JASIS*, 45, 10, pp. 745-752.

MACLEOD, I.A. (1990): "Storage and Retrieval of Structured Documents." *Information Processing & Management*, 26, 2, pp. 197-208

MACLEOD, I.A. (1991): "Text retrieval and the relational model." *JASIS*, 42, 3, 155-165.

MANDELBAUM, J.B. (1992): *Small Automation Projects for Libraries and Information Centers.* Wesport: Meckler

MARTIN, J. (1990): *Hyperdocuments & How to Create Them.* Englewood Cliffs: Prentice-Hall.

MCKNIGHT, C., DILLON, A. y RICHARDSON, J. (1991): *Hypertext in Context*. Cambridge: Cambridge University Press.

MEADOW, C.T. (1992): *Text Information Retrieval Systems*. San Diego: Academic Press.

MEGHANI, C., RABITTI, F. y THANOS, C. (1991): "Conceptual Modeling of Multimedia Documents." *Computer*, October, pp. 23-30.

MICHAEL, D.A. (1994): "What is Used during Cognitive Processing in Information Retrieval and Library Searching?" *JASIS*, 45, 7, pp. 498-514.

MIGUEL, A. DE, y PIATTINI, M. (1993): *Concepción y diseño de bases de datos*. Madrid: RA-MA

MIGUEL, A. DE, y PIATTINI, M. (1995): "Visión general de los repositorios y diccionarios." Piattini, M y Daryanani, S.N. (eds*.) Elementos y herramientas en el desarrollo de sistemas de información*. Madrid: Ra-Ma, pp.229-239.

MORRIS, A.(ed) (1992): *The Application of Expert Systems in Libraries and Information Centres.* Bowker-Saur.

MOTA, L., CELMA, M., y CASAMAYOR, J.C. (1994): *Bases de datos relacionales: teoría y diseño*. Valencia: Servicio de Publicaciones U.P. Valencia.

MOYA ANEGÓN, F. DE (1994): *Los sistemas integrados de gestión bibliotecaria*. Madrid: ANABAD.

NEILL, S.D. (1992): *Dilemmas in the Study of Information. Exploring the Boundaries of Information Science*. New York: Greenwood Press.

NICHOLLS, P. y RIDLEY, J. (1996): "A Context for Evaluating Multimedia." *Computers in Libraries*, 16, 4, pp. 34-39.

NIELSEN, J. (1990): *Hypertext and Hypermedia*. San Diego: Academic Press.

ORERA ORERA, L.(ed.) (1996): *Manual de Biblioteconomía*. Madrid: Síntesis.

ORTIN-REPISO, V. y OLMEDA, C. (1994): "Orientarse en Internet: Hytelnet 6.4" *Revista Española de Documentación Científica*, 17, 1, pp.9-24.

OZKARAHAN, E. (1995): "Multimedia Document Retrieval*." Information Processing & Management*, 31, 1, pp.113ss.

PAICE, C. D. (1991): "A Thesaural Model of Information retrieval*." Information Processing & Management*, 25, 5, pp. 433-447.

PARSAYE, K. *et alii* (1989): *Intelligent databases : object-oriented, deductive hypermedia technologies*. New York : John Wiley & Sons.

PEIS, E. y FERNÁNDEZ-MOLINA, C. (1994): "Evaluación de la recuperación de información por materias en los OPACs." *Actas de las IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94*, Gijón: Universidad de Oviedo, pp.245-251.

PRAX, J-Y. (1994): *La gestion électronique documentaire*. Paris: Armand Colin.

RECODER, M.J., ABADAL, E., y CODINA, L. (1991): *Información electrónica y nuevas tecnologías*. Barcelona: PPU.

REID, C. (1990): "Comparing text, document and relational database management systems." *Library Software Review*, Marzo-Abril, pp. 80-82.

REINGRUBER. M.C., GREGORY, W.W (1994): *The Data Modelling Handbook*. New York: John Wiley & Sons

REYNOLDS, D. (1989): *Automatización de biblioteca*s. Madrid: Fundación Germán Sánchez Ruipérez.

RIJSBERGEN, C.J. VAN (1979): *Information Retrieval*. London: Butterworths.

RIJSBERGEN, C.J. VAN y LALMAS, M. (1996): "Information Calculus for Information Retrieval." *JASIS*, 47, 5, pp.385-398.

ROBERTSON, S.E. (1977): "Theories and models in information retrieval*." Journal of Documentation*, 33, 2, pp. 26-148

RODGERS, U. (1990): *UNIX Database Management Systems*. Englewood Cliffs: Yourdon Press.

RODRÍGUEZ DE LAS HERAS, A. (1991): *Navegar por la información*. Madrid: Fundesco.

ROWLEY, J.E. (1992*): Organizing Knowledge: An Introduction to Information Retrieval*. Aldershot: Ashgate.

ROWLEY, J.E. (1994): *The Basics of System Analysis and Design for Information Managers.* London: Clive Bingley.

SAFFADY, W. (1986): *Informática documental para bibliotecas*. Madrid: Díaz de Santos.

SAFFADY, W. (1989): *Text storage and retrieval systems*. London: Mekler.

SALTON, G. (1989): *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Reading: Addison-Wesley.

SALTON, G., y BUCKLEY, C. (1990): “Improving Retrieval Performance by Relevance Feedback.” *JASIS*, 41, 4, pp. 288-297

SALTON, G. y BUCKLEY, C. (1994): “Automatic Structuring and Retrieval of Large text files.” *Communication of the ACM*, 37, 2, pp. 97-108.

SALTON, G., y MCGILL, J.M. (1983): *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

SARACEVIC, T. *et alii* (1988): “A study of information seeking and retrieval.” *JASIS*, 39, pp. 161-216

SATOR, F. (1992): “Acceso integrado a bases de datos heterogéneas.” en MAP, *Encuentro sobre bases de datos en la Administración Pública*, Madrid, pp.195-200.

SCHAMBER, L. (1996): “What Is a Document? Rethinking the Concept in Uneasy Times.” *JASIS*, 47, 9, pp. 669-671.

SCOTT BRAND, P. (1996): “Relevancy and Searching the Internet.” *Computers in Libraries*, 16, 8, pp.35-48.

SEYER, P. (1991): *Hypertext. Concepts and Applications*. Blue Ridge Summit: Windcrest Books.

SIEWERT, M.E.C. (1996): “Full-Text Information Retrieval: Introduction.” *JASIS*, 47, 5, pp. 261-262.

SOCADI, (1992): *Formatos bibliográficos: su compatibilidad y conversión. Casos de usuarios de sistemas automatizados de bibliotecas*. Barcelona: Societat Catalana de Documentació i Informació.

SOERGEL, D. (1985): *Organizing Information. Principles of Data Base and Retrieval Systems*. New York: Academic Press.

STEWART, L. (1990): *Public access CD-ROMs in libraries: case studies*. London: Meckler.

TAGUE-SUTCLIFFE, J. (1994): “The pragmatics on information retrieval experimentation, revisited.” *Information Processing and Management*, 28, 4, pp. 467-490.

TAGUE-SUTCLIFFE, J. (1996): “Some Perspectives on the Evaluation of Information Retrieval Systems.” *JASIS*, 47, 1, pp. 1-3.

TARE, R.S. (1990): *Procesamiento de datos en UNIX*. Madrid: McGraw-Hill.

TEDD, L. A. (1988): *Introducción a los sistemas automatizados de bibliotecas*. Madrid: Díaz de Santos.

TOFFLER, A. (1993): *La Tercera Ola.* Barcelona: Plaza y Janés.

TRAMULLAS SAZ, J. (1995): “Una introducción a la Informática Documental.” *Apuntes del Centro de Cálculo de la Universidad de Zaragoza*, 8, pp. 6-10.

TRAMULLAS SAZ, J. (1996a): *Apuntes de Informática Documental*. Zaragoza: Kronos.

TRAMULLAS SAZ, J. (1996b): “Investigación en sistemas bibliográficos personales: una revisión de aplicaciones *freeware y shareware*.” Tramullas, J. (Ed.) *Tendencias de investigación en Documentación*. Zaragoza: Universidad de Zaragoza, pp.51-73.

TRAMULLAS SAZ, J. (en prensa): “Recuperación de información en WWW: planteamiento, herramientas y perspectivas.” *Anuari de SOCADI*, Barcelona: SOCADI.

TRAMULLAS SAZ, J. (sin publicar): “Internet e Intranet: la creación de servicios de información en la organizaciones*.” III Jornadas Nacionales de Información y Documentación Empresarial*, Murcia, 1996.

TRAMULLAS SAZ, J., y CUBILLO, A. (1995): *BRS/Search: Introducción a los sistemas de gestión de bases de datos documentales*. Zaragoza: Kronos.

TSENG, G., POULTER, A. y HIOM, D. (1995): *The Library and Information Professional’s Guide to the Internet*. London: Library Association Pub.
TURTLE, H.R. y CROFT, W.B. (1992): “A Comparison of Text Retrieval Models*.” The Computer Journal* , 35, 3, pp. 279-290.

TURTLE, H. y FLOOD, J. (1995): “Query Evaluation: Strategies and Optimizations.” *Information Processing & Management*, 31, 6, pp. 831-850.

UBIETO, A.P. (1994): “Herramientas telemáticas en Internet para la Documentación*.” Actas de las IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94*, Gijón: Universidad de Oviedo, pp. 57-62.

UBIETO, A.P. (1995*): Documentación Automatizada. Manual de uso de la red Internet*. Zaragoza: Anubar.

VICKERY, B.C., y VICKERY, A. (1992): *Information Science in Theory and Practice.* London: Bowker-Saur.

WALKER, G. y JONES, J. (1993*): Online Retrieval: A Dialogue of Theory and Practice*. Englewood: Libraries Unlimited.

WARNER, J. (1990): "Semiotics, Information Science, Documents and Computers." *Journal of Documentation*, 46, 1, pp. 24ss.

WATTERS, C.R. (1989): “Logic Framework for Information Retrieval.” *JASIS*, 40, 5, pp. 311-324.

WEITZMAN, E.A., y MATTHEW, B.M. (1995): *Computer Programs for Qualitative Analysis*. Thousand Oaks: SAGE Publications, Inc.

WILLET, P. (ed.) (1988): *Document Retrieval Systems (The Foundations of Information Science, vol. 3)*. London: Taylor Graham.

WORMELL, I. (ed.) (1988): *Knowledge Engineering: Expert Systems and Information Retrieval.* London: Taylor Graham.

WRIGHT, K.C. (1995): *Computer Related Technologies in Library Operations*. Aldershot: Gower.

Página en blanco

La primera edición de *Introducción a la Documática. 1. Teoría*, se redactó en 1996. En los catorce años que han transcurrido desde entonces, la evolución y la integración de las aplicaciones tecnológicas en el trabajo informativo y documental ha alcanzado tal importancia, que el mismo no puede entenderse ni realizarse adecuadamente sin ellas. La presente reedición digital es la última que va a realizarse de forma cerrada. Desde 2010, la obra se pasa a publicar bajo licencia GNU Public Documentaion License, con la esperanza de que pueda ser actualizada y renovada por los usuarios interesados.

Jesús Tramullas es profesor titular en el Departamento de Ciencias de la Documentación de la Universidad de Zaragoza. Autor de más de un centenar de trabajos sobre diferentes aspectos de la aplicación de herramientas tecnológicas en la gestión de información, refleja su actividad en su web personal,

http://tramullas.com,

"Operaba en un estado adrenalítico alto y casi permanente, conectado a una consola de ciberespacio hecha por encargo que proyectaba su incorpórea conciencia en la alucinación consensual que era la matriz."

W. Gibson, *Neuromante*.

